2006

Controlled integration of the Ty5 retrotransposon in Saccharomyces ceverisiae [i.e. cerevisiae]

Junbiao Dai
Iowa State University

Controlled integration of the Ty5 retrotransposon in *Saccharomyces ceverisiae*

by

Junbiao Dai

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Molecular, Cellular and Developmental Biology

Program of Study Committee:
Daniel F. Voytas, Major Professor
Linda Ambrosio
Janice Buss
Michael Shogren-Knaak
Yanhai Yin

Iowa State University

Ames, Iowa

2006

Graduate College
Iowa State University

This is to certify that the doctoral dissertation of

Junbiao Dai

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For the Major Program

TABLE OF CONTENTS

ABSTRACT

One essential step in the life cycle of retroelements is the stable integration of a copy of retroelement cDNA into the host genome. Random integration is potentially hazardous and could have deleterious genetic effects to the host. Therefore, elements and their hosts have coevolved mechanisms to regulate retroelement integration. In the budding yeast *Saccharomyces cerevisiae*, the Ty5 retrotransposon preferentially integrates into domains of heterochromatin. Targeting to these locations is determined by interactions between an amino acid sequence motif at the C-terminus of Ty5 integrase (the targeting domain) and the heterochromatin protein Sir4p. New Ty5 integration hotspots are created when Sir4p is tethered to ectopic DNA sites. Targeting to sites of tethered Sir4p is abrogated by single amino acid substitutions in either the targeting domain or Sir4p that prevent their interaction. Ty5 target specificity can be altered by replacing the integrase targeting domain with other peptide motifs that interact with known protein partners. Integration occurs at high efficiency and in close proximity to DNA sites where the protein partners were tethered. These findings indicate that Ty5 actively selects integration sites, and that targeting determinants are modular. These findings also suggest ways in which retroviral integration can be controlled, namely through addition of peptide motifs to viral integrases that direct integration complexes to specific chromosomal sites.

The targeting domain of Ty5 integrase is post-translationally modified *in vivo*. Analysis by tandem mass spectrometry analysis revealed that the second serine within the targeting domain (S1095) is phosphorylated. Phosphorylation of the S1095 is required for targeted integration as measured by a plasmid-based targeting assay. Using surface plasmon resonance spectroscopy, S1095 phosphorylation was found to be required for productive interaction with Sir4C *in vitro*. This provides direct evidence for the requirement of post-translational modification in Ty5 targeting and reveals one mechanism that cells have adopted to overcome the deleterious effect of transposable element invasion; that is, they control integration specificity by modifying element-encoded proteins. Prompted by the discovery of targeting domain phosphorylation, a total of 109 kinase deletion strains were screened for kinases that affect Ty5 target specificity. Among eight candidates, the DNA

damage checkpoint kinase Dun1p was found to be required, either directly or indirectly, for Ty5 transposition. The role of kinases in targeting was further explored by identifying which of the 109 non-essential protein kinases in *S. cerevisiae* affect the integrity of heterochromatin. Using transcriptional silencing as a measure of heterochromatin integrity, several kinases, including Dun1p, were identified that affect transcriptional silencing when mutated, principally at telomeres. Interestingly, most kinases are members of MAP pathways involved in DNA damage, osmolarity, cell wall integrity and pheromone responses. These results suggest that integrity of heterochromatin is tightly controlled by protein kinase cascades, which may indirectly contribute to the regulation of Ty5 integration specificity under different stresses.

To determine which kinases directly phosphorylate Ty5 integrase, a portion of the integrase C-terminus (mINC) was purified and used as a substrate for *in vitro* kinase assays. All of the kinases encoded by the yeast genome (125 in total) were purified as GST-fusion proteins and tested their ability to phosphorylate mINC. Sixteen kinases were identified that modify mINC, and deletions in genes encoding these kinases affect Ty5 integration specificity. Interestingly, mutations in two kinases (*ime2*Δ and *rck2*Δ) increased Ty5 integration specificity significantly. The remainder impaired integration specificity as measured by the plasmid-based targeting assay. Using several different mINC constructs with various serine/threonine mutations, the kinases were grouped based on their ability to phosphorylate a defined set of serine/threonine residues. Four kinases – Hrr25p, Rim11p, Rck2p and Yak1p – are the most likely candidates for phosphorylating S1095. Phosphorylation of mINC by multiple kinases, and importantly, the observation that mutations in these kinases affect Ty5 integration specificity, suggests that both transposition and integration of Ty5 are directly regulated in response to cellular processes or environmental stress.

CHAPTER 1. GENERAL INTRODUCTION

TRANSPOSABLE ELEMENTS: DISCOVERY AND CLASSIFICATION

The information to build, control and maintain a living organism lies in its chromosomes. A chromosome is a compact macromolecule composed of a very long, continuous piece of DNA and associated proteins. The genetic information is stored as units in DNA segments called genes, which can be transcribed into RNA and translated into proteins. The genes are often fixed at a defined chromosomal location, although in some rare conditions they can be inverted or translocated. However, some of the DNA segments are able to move more frequently from one position to another. Those segments are called transposable elements (TEs) or mobile DNA.

The concept of mobile DNA was actually first documented prior to the discovery of the structure of DNA. In the 1940's, the maize cytogeneticist, Barbara McClintock, made a series of observations that the so-called "controlling elements", which she referred to as *Dissociator(Ds)* and *Activator(Ac),* were able to move or change position on the chromosomes (Mc, 1950). This conflicted with the common concept that the genetic material is static and unchanging between generations. The significance of her work about TEs was not recognized by the scientific community until the late 1960s and early 1970s, after the discovery of transposition in bacteria and yeast. Today, of course, TEs have been identified in all organisms that have been investigated, and in some organisms like maize, they make up over 50% of the genomic DNA (SanMiguel et al., 1996).

TEs are usually divided into two classes: retrotransposons and DNA transposons. The retrotransposons will be discussed in detail in the next section. First characterized in maize, the DNA transposons have been found to exist in most organisms and are the major group of TEs in prokaryotes. These elements have a size ranging from several hundred bases to about 10 kb, and they are flanked by terminal inverted repeats (TIRs) (Figure 1). TIRs can be as short as 11bp (*Ac/Ds*) or as long as a few hundred bases (*Mutator*). DNA transposons can be divided into different families. In each family, at least one member encodes an enzyme named transposase, which will catalyze the movement of that family of transposons by

recognizing the specific TIRs. Families, therefore, are defined by the sequences of their transposases and TIRs. Elements capable of transposition are autonomous members of the family (like *Ac* in the *Ac/Ds* family). Other members, in which the transposase is inactivated by mutations, are called non-autonomous elements (like *Ds*). Their transposition relies on the autonomous elements.

During transposition, the transposase recognizes and binds the TIRs, and then excises the element at the termini from the donor DNA. The gap left by the excision of the element is repaired, either by ligation of the two broken ends or by recombination. The transposase will also make nicks on the target DNA, which separates the target DNA and generates two 5' overheads. The element is inserted between the break site, leaving a single strand gap at both ends. The gaps are then filled by host DNA repair machinery, which generates the target site duplication (TSD), a hallmark of transposition.

The above mechanism is called "cut-and-paste" transposition, which will normally only move the element from one chromosomal location to another, but won't increase its copy number. However, different transposons have evolved different mechanisms to amplify themselves. The excision of the *Ac* element is activated during or after it is replicated in S phase. The excised elements preferentially insert into the region where replication has not taken place, which results in three copies of *Ac* by the end of S phase. Many other transposons, like the P elements in *Drosophila* (Engels et al., 1990), are restored at their original locus after excision by recombination. In these instances, the other chromosome or chromatid, which does not undergo excision, is used as a template for repair.

RETROTRANSPOSONS: GENOME STRUCTURE AND LIFE CYCLE

In eukaryotes, the most abundant TEs are retrotransposons. Unlike transposons, retrotransposons transpose via a RNA intermediate. The retrotransposons can be divided into LTR and non-LTR retrotransposons based on whether they are flanked by long terminal repeats (LTRs). The LINE-1 (L1) elements are the most abundant autonomous non-LTR retrotransposon in human, composing about 17% of the human genome. A full-length L1 element contains a 5' UTR, followed by two long open read frames (ORF1 and ORF2), then

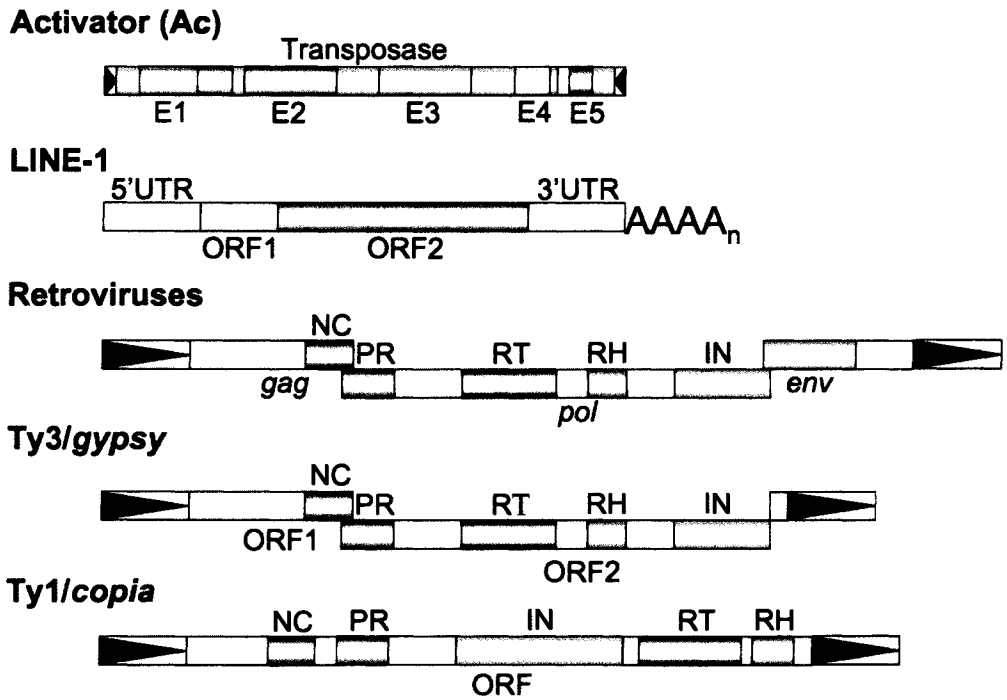

Figure 1. Schematic genome structure of transposable elements. Activator (Ac) represents DNA transposons; LINE-1 represents non-LTR retrotransposons. Black triangle represents the long terminal repeats (LTRs). Red triangle represents the inverted repeats (IRs). E1-E5, Exon 1-Exon5; UTR, untranslated region; NC, nuclear capsid; PR, protease; RT, reverse transcriptase; RH, RNase H; IN, integrase; env, envelope gene.

a short 3' UTR and the poly (A) tail (Figure 1). ORF2 encodes a polymerase (Lander et al.) that has functional enzymes like reverse transcriptase and AP endonuclease (Han et al., 2004). The endonuclease creates a nick on host DNA, and reverse transcriptase copies the L1 mRNA into cDNA using the 3' end of the nicked host DNA as a primer. This mechanism is called target-primed reverse transcription (Luan et al., 1993).

The LTR retrotransposons are typically larger than the non-LTR elements and range in size from several hundred bases to more than 10kb. The size of LTR also varies in length from several hundred to several thousand bases. LTR-retrotransposons are found in all eukaryotes,

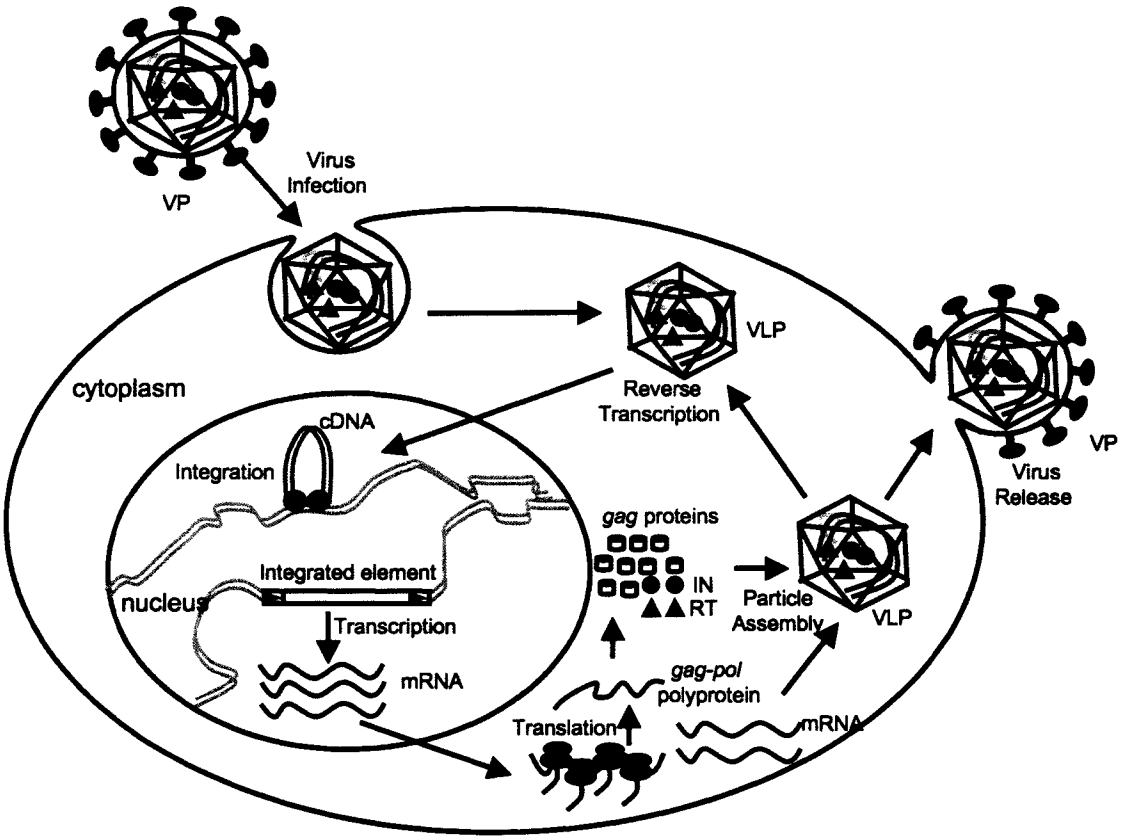

Figure 2. Life cycle of LTR retroelements (retrotransposons and retroviruses). IN, integrase; RT, reverse transcriptase; VLP, virus-like particle; VP, virus particle. Blue triangle represents LTRs.

probably indicative of an early origin. The internal region of LTR-retrotransposons usually contains two major open reading frames that encode the Gag protein and the *Pol* proteins (berg and howe 1989). The Gag protein is a structural protein that assembles into virus-like particles (VLPs) (see below), whereas the Pol protein contains several enzymatic proteins such as protease (PR), reverse transcriptase (RT) and integrase (IN) (Figure 1). There are two major sub-classes of LTR-retrotransposons called *gypsy* and *copia*, differing in the order of RT and IN within the Pol protein (Figure 1). Retroviruses are similar to *gypsy* elements, except they encode an extra open reading frame called the envelope (*env*) gene (Figure 1). The acquisition of *env* enables retroviruses to be released from one cell and infect other cells or organisms (Figure 2). So far, the retroviruses are only identified in animals, although in plants, some LTR-retrotransposons also contain an open reading frame at the site

corresponding to the *env* gene (Peterson-Burch et al., 2000; Wright and Voytas, 1998). The function of these genes in the plant elements is unknown.

The life cycle of the LTR retrotransposons can be divided into several major steps: transcription, particle formation, reverse transcription and integration (Figure 2). LTR retrotransposons are first be transcribed into mRNA by the pol II transcription machinery. The mRNA is exported into the cytoplasm where it will either be used as template for protein synthesis or, along with some of element-encoded proteins, is packaged into VLPs. Within the VLP, the mRNA is reverse transcribed to a linear, double-stranded cDNA by the element-encoded RT using a tRNA as the primer. In association with IN and possibly other element- or host-encoded factors, the cDNA is imported into the nucleus as a pre-integration complex (Spicuglia et al.). The cDNA is then integrated into the genome, a step carried out by IN. Similar to DNA transposons, LTR-retrotransposon integration leaves a 4-6 bp TSD.

TES AND GENOME EVOLUTION

Ubiquitous distribution of TEs.

One common finding of the genome-sequencing projects is that the majority of genomes are composed of non-coding DNA, and the amount of non-coding DNA is proportional to genome size. Other than some regulatory elements and simple repetitive DNA, the largest amount of non-coding DNA is made of TEs. For example, in humans, 50% or more of the genome is comprised of repeat sequences, whereas the coding region is less than 5% (Lander et al., 2001). Among the repeat sequences, about 90% are derived from TEs. All three classes of TEs are present in the human genome. The DNA transposons make up about 2 to 3% of the genome, which are mainly MER1 and MER2 elements (Roy-Engel mobile DNA II). The LTR retrotransposons make up 5 to 10% of the human genome, including MaLR, Mer4 and a large amount of endogenous retroviruses. Almost all of the DNA transposons and LTR retrotransposons in human are non-functional. The majority of the TEs in humans are non-LTR retrotransposons, including LINEs and SINEs. The L1 elements are the major form of the currently active elements.

To date, several plant genomes have been sequenced including *Arabidopsis* and rice.

Extensive investigations have been carried out to analyze the repeat DNA sequences in maize (Messing et al., 2004; Meyers et al., 2001; SanMiguel et al., 1996), wheat (Li et al., 2004; Sabot et al., 2005) and barley (Kalendar et al., 2004; Schulman and Kalendar, 2005). One interesting outcome of these studies is that genome size of many plant species differs as a result of variable amounts of LTR-retrotransposons integrated into the intergenic regions. In plants with small genomes such as *Arabidopsis* (125 Mbp), only a small portion (perhaps less than 5%) of the genome is comprised of TEs, and the TEs are clustered in the pericentromeric regions (Initiative, 2000). In maize, which has a moderately large genome size of ~2500 Mbp, at least about 60% of the genome is composed of TEs (Messing et al., 2004; Meyers et al., 2001). The LTR retrotransposons are the predominant TEs in maize, comprising 56% of the genome, whereas the DNA transposons make up only about 1% (Messing et al., 2004). In contrast to the human genome, which contains relatively large amounts (> 25%) of non-LTR retrotransposons (Lander et al., 2001), there is only very small amount (< 0.2%) of non-LTR retroelements in the plant genome (Messing et al., 2004).

Besides human and plant genomes, TEs have been identified in every organism investigated. Given the abundance of elements in the genome, it is tempting to speculate that they have long-term effects on the structure and function of the genome. Actually, the view of TEs as 'selfish DNA" or "junk DNA" is changing as the effects of TEs on genomes are increasingly elucidated.

TEs and telomere maintenance.

Unlike bacteria, which contain a circular chromosome, most eukaryotic genomes are comprised of linear chromosomes. The chromosome ends are capped with repeated DNA arrays called telomeric repeats. The sequence and length of the telomeric repeats differs from species to species, but they are usually between 6 to 10 bp long and are G/T rich. The telomeric repeats, along with the various unique DNA binding proteins, form the telomere, which prevents chromosome ends from behaving like broken DNA and fusing. Typically, telomeres are maintained during the cell cycle by an enzyme call telomerase, which adds extra telomeric repeats by reverse transcribing an RNA template included in the telomerase holoenzyme (Blackburn, 1992). Lack of telomerase causes telomere shortening. As a result,

telomeres gradually shorten to a point where they initiate senescence and eventually lead to chromosome fusion and cell death (Harley et al., 1990).

Interestingly, *Drosophila* has no telomerase and lacks short telomeric repeats. Instead, Drosophila telomeres are composed of multiple copies of two telomeric specific retrotransposable elements, *Het-A* and *TART* (reviewed by (Pardue et al., 2005)). Both are non-LTR retrotransposons with lengths of ~6 kb and 10 kb, respectively. *Het-A* and *TART* transpose specifically to chromosomal ends and form long head-to-tail arrays. These TE-comprised telomeres, like other eukaryotic telomeres, are dynamic structures. Although they share some common features with other non-LTR retrotransposons, *Het-A* and *TART* are special to *Drosophila*. They contain a significant portion of non-coding DNA, and they don't require nicked DNA for integration, presumably because they integrate onto the chromosome ends. Their Gag proteins efficiently localize to nuclei at well-defined sites, whereas the Gag proteins of other non-LTR retrotransposons remain predominantly in the cytoplasm. *TART* Gag is recruited to chromosomal ends by *Het-A* Gag (Pardue et al., 2005).

Although *Drosophila* is the only organism known whose telomere is comprised of retrotransposons instead of telomeric repeats, several other organisms have TEs associated with telomeres. For example, in yeast, the Y' elements form tandem repeats of up to four head-to-tail copies on some chromosomes. When telomerase is inactivated, survivors that resume growing have variably amplified arrays of Y' elements, indicating the Y' element may participate in telomere maintenance in the absence of telomerase (Maxwell et al., 2004). The other yeast retrotransposon, Ty5, can also be found preferentially at telomeres and silent mating loci (see also the discussion below) (Zou et al., 1996a). In silkworms (Bombyx mori), at least two non-LTR retrotransposons (TRAS1 and SART1) integrate specifically into the TTAGG telomerase repeats (Takahashi and Fujiwara, 1999).

TEs and centromere evolution.

Several lines of evidence suggest TEs may be involved in the evolution of centromeres. First, centromere-specific TEs have been identified in multiple organisms, including different grass species such as maize, rice, barley, wheat and rye (Jiang et al., 2003;

Langdon et al., 2000; Nagaki et al., 2003; Wu-Scharf et al., 2000). The centromere-specific TEs in maize (CRM) are exclusively interspersed within the 156 bp-centromeric-satellite repeats (CentC) (Jiang et al., 2003; Nagaki et al., 2003). Similarly, a centromere-specific TE in rice (CRR) is found specifically within the 155bp CentO repeats (Nagaki et al., 2004). In the above examples, the centromere-specific TEs may integrate specifically into centromeric repeats. In addition, numerous examples indicate that centromeric repeat arrays have evolved from TEs. For example, the centromeric tandem repeat of the wheat species *Aegilops speltoides* shares high similarity with part of *cereba*, a Ty3/*gyspy*-like retrotransposon (Irwin et al., 2005). Similarly, the satellite arrays in pericentromeric regions in *Arabidopsis* may have evolved from one of the En/Spm-like families (Atenspm) (Kapitonov and Jurka, 1999). Finally, there is a link between TEs and heterochromatin formation. The highly conserved centromere-associated protein, CENP-B, has extensive sequence similarity with the transposase proteins encoded by the Tc1/mariner class of transposase (Smit and Riggs, 1996), suggesting a close relationship between CENP-B and transposase. CENP-B homologs in *S. pombe*, Abp1, Cbh1 and Cbh2, also specifically bind to the centromeric repeats dg and dh. Deletion of these proteins impairs initiation of heterochromatin formation by diminishing centromeric association of Swi6 and decreasing H3 K9 methylation (Nakagawa et al., 2002). These transposase-originated proteins, therefore, may act as site-specific nucleation factors for the formation of centromeric heterochromatin.

TEs and chromosomal rearrangements.

TEs are able to reshape the host genome by inducing various chromosomal rearrangements such as deletions, duplications and inversions. Two possible mechanisms have been described (reviewed by(Gray, 2000)) that involve either an indirect role of TEs in homologous recombination or a direct role in alternative transposition. In yeast, normally the Ty elements (mainly Ty1) are recombination cold spots, despite their high copy number in the genome (32-full length Ty1 elements (Kim et al., 1998)). However, when yeast is treated with DNA damaging agents such as UV irradiation, the recombination frequency increases dramatically (Kupiec and Steinlauf, 1997). Elegant work done in Petes' lab has shown that the head-to-head Ty1 array on chromosome III is one of the preferred fragile sites of double strand breaks when DNA replication is compromised by low levels of DNA polymerase

(Lemoine et al., 2005). They further showed that when the cells are exposed to UV irradiation, the chromosomal sites where Ty1 resides become double strand break hotspots and cause various chromosomal rearrangements including translocations and duplications (personal communication).

Alternative transposition can have a large impact on chromosome organization. Alternative transposition occurs when the transposase of a DNA transposon binds to the ends of two different elements, sometimes elements on different chromosomes or sister chromatids. If transposase cleaves these elements, the result often leads to chromosome breakage and various types of chromosome rearrangements (Zhang and Peterson, 1999; Zhang and Peterson, 2005).

V(D)J recombination

The diversity of antigen receptor repertoires is accomplished by assembling receptor genes from multiple variable (V), joining (J) and sometimes diversity (D) gene segments, a process called V(D)J recombination. Two genes, recombination-activating gene 1 (RAG1) and RAG2, initiate the process by introducing double strand breaks at the target gene segments. The broken ends are rejoined by DNA repair factors as well as the RAG complex, generating variability from imprecise joining of the coding sequences. The RAG1 and RAG2 genes derive from transposons and have been co-opted by hosts to create a transposase-like enzyme. Detailed mechanisms of V(D)J recombination have been reviewed recently (Schatz and Spanopoulou, 2005; Spicuglia et al., 2006).

TEs and gene evolution

In addition to the above-mentioned RAG genes and *Het-A* and *TART* elements , increasing evidence indicates that TEs have contributed significantly to the diversity of cellular genes. Based on similarity between gene coding sequences and TEs, independent studies have identified 19 (Smit, 1999) and 47 (Lander et al., 2001) human genes that appear to have coding regions derived in part from TEs. Another study used REPEATMASKER software to identify 533 cases of TE sequences within human protein coding sequences (Nekrutenko and Li, 2001). More recently, using amino acid sequence probes translated from

TEs in all reading frames (ignoring stop codons), Britten identified 1950 different human proteins, which might derived from TEs (Britten, 2006). It is possible that the actual number of human proteins containing TE sequences is even larger, but due to sequence drift, they are difficult to recognize. Wheelan et al. identified three human genes, which seem to be broken by L1 element insertions (Wheelan et al., 2005). These genes are split into two smaller transcripts, suggesting a novel method to generate new functional genes by TEs. In plants, a recently identified class of TEs called Helitrons and a form of DNA transposon called Pack-MULEs were shown to acquire and fuse fragments of plant genes (reviewed by (Bennetzen, 2005)). Transcription and translation of these gene fragments may provide raw material for the evolution of novel genes and novel genic functions. Bejerano, Lowe et al. reported that a distal enhancer and an ultraconserved exon are derived from a previous unknown SINE retroposon (Bejerano et al., 2006). The list of examples in which TEs have acquired functions to serve their host will keep growing, and more novel functions of TEs in their host cells will undoubtedly be revealed in the future as additional genomes are characterized.

CONTROL OF TE ACTIVITY

Since TEs can act as mutagens by insertional mutagenesis or activation or repression of adjacent genes, under normal conditions, their activity within the host must be tightly controlled. On the other hand, during some circumstances, for example, when cells are facing dramatic environmental changes and are required to evolve quickly, the TEs could be activated, which in turn, will provide the host with genotypic diversity that will increase its ability to survive the new environment.

TEs can be passively inactivated. When TEs invade a host genome, integration of the elements may have a deleterious effect, resulting in the death of host cell. Among the host cells who survive the TE invastion, the elements are not under positive selection. Mutations will quickly accumulate within the TE coding sequence, which inactivates the proteins and leads to degenerate or deleted elements. This process is called "vertical inactivation"(Lohe et al., 1995). The mutations within an element are proportional to the time of its transmission into the genome. This process will make most of the TEs within the genome non-functional, and they become the "junk DNA" that characterizes a significant fraction of most genomes.

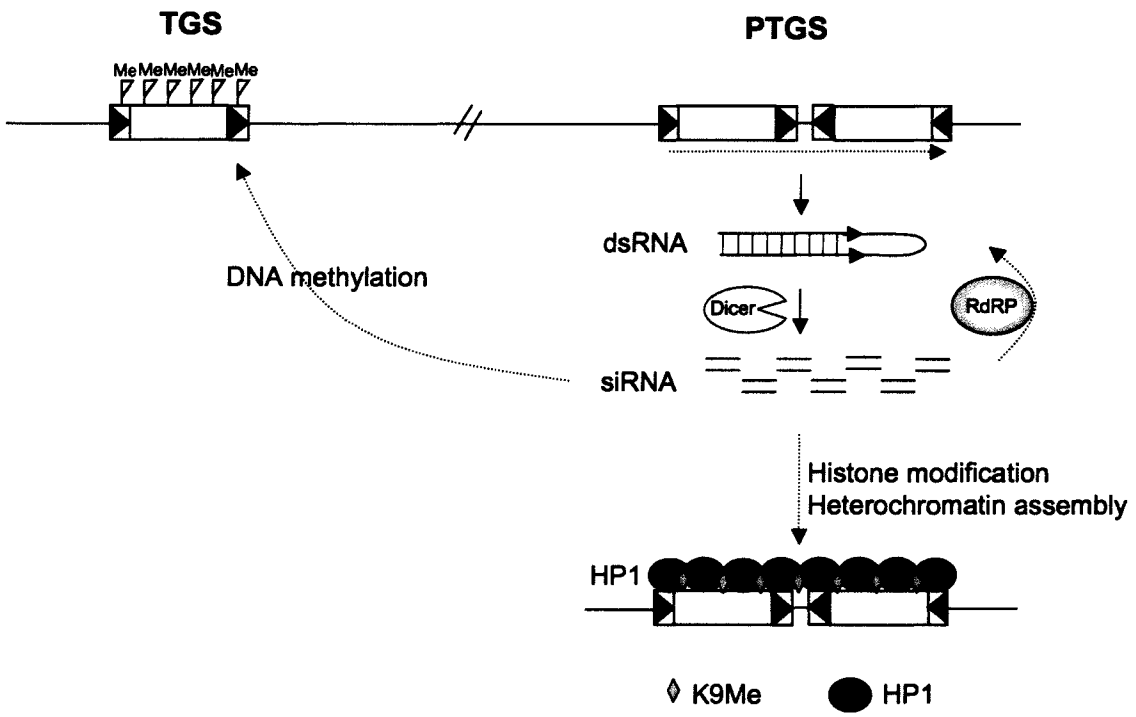

Figure 3. The repression of transposable elements by both transcriptional gene silencing (TGS) and post-transcriptional gene silencing (PTGS). Redrawn and adapted from Wong & Choo, *Trends Genet.*, 2004, 20(12):611-6.

At least two mechanisms are known to actively silence intact TEs (Figure 3): (1) transcriptional gene silencing (TGS) and (2) post-transcriptional gene silencing (PTGS). TGS is generally associated with DNA methylation. DNA methylation is considered a primary defense mechanism against TEs in both plants and animals (Yoder et al., 1997). In *Arabidopsis*, CG methylation (carried out by DNA methyltransferase MET1) and non-CG methylation (carried out by DNA methyltransferase CMT3) function redundantly in the immobilization of TEs. High-frequency transposition of CACAT1, an endogenous DNA transposon, was only observed in the *met1, cmt3* double mutant (Kato et al., 2003). CACAT1 is also mobilized in a *ddm1* mutant, in which all DNA methylation is abolished (Miura et al., 2001). The activated CACAT1 in the *ddm1* mutant remains active in the presence of a wild type DDM1 gene, suggesting re-establishment of DNA methylation is not sufficient to

silence the element (Kato et al., 2004). Other epigenetic marker(s) may be critical for its mobility. Histone H3 K9 methylation may guide methylation of TEs (Gendrel et al., 2002). In Neurospora, a HP1 (heterochromatin protein 1) homolog is localized to heterochromatin foci and required for DNA methylation, suggesting it may also play a role in silencing TEs (Freitag et al., 2004).

RNAi-mediated gene silencing is considered a form of cellular immunity against viruses and repeated genome sequences (reviewed by (Buchon and Vaury, 2006)). It is initially triggered through the synthesis of double-strand RNA (dsRNA), a product of RNA-dependent RNA polymerase (*RdRP*). The dsRNA is recognized and cleaved by the RNase III helicase, Dicer (*Dcr*), generating small interfering RNAs (siRNAs) of 21-25 nucleotides. siRNAs are incorporated into the RNA-induced silencing complex (RISC), in which they guide RISC to degrade the homologous mRNAs. The siRNAs can also direct the sequence-specific modification of DNA and chromatin (Figure 3). Mutations in genes in the RNAi pathway result in mobilization of families of transposable elements in *D. melanogaster* (Aravin et al., 2004), in *Arabidopsis* (Zilberman et al., 2003), in *C. elegans* (Ketting et al., 1999; Tijsterman et al., 2002) and in many other species (Chicas et al., 2004; Wu-Scharf et al., 2000). Recently, Slotkin et al. cloned and characterized a naturally occurring locus able to heritably silence the otherwise highly active MuDR transposon in maize (Slotkin et al., 2005). This locus, named Mu killer (Muk), results from the inverted duplication of a partially deleted autonomous MuDR element located at the breakpoint of a genomic deletion. Muk produces a hybrid hairpin transcript that is processed into small RNAs, which are amplified when the target MuDR transcript is present. The MuK provides the first example of a naturally occurring transposon derivative capable of initiating the heritable silencing of an active transposon family.

As discussed above, the activities of TEs are under various controls by the host defense systems. However, in some stress conditions that reduce host fitness, TEs can also be activated. In the *Solanaceae* plant family, the Tnt1 elements (a superfamily of LTR-retrotransposons) are activated only in response to stress and microbial factors (reviewed by (Grandbastien et al., 2005)). Tnt1 retrotransposons have a high level of variability in their

LTR sequences, which direct transcription. The LTRs have evolved to confer new expression patterns, mostly associated with responses to diverse stress conditions. The numbers of Tnt1 retrotransposons in different *Solanaceae* species, therefore, probably correlates with different expression conditions that relate to the environmental history of each host.

More detailed studies on the activation of TEs have been carried out in *S. cerevisiae* (recently reviewed by (Lesage and Todeschini, 2005)). Of the five LTR retrotransposons in yeast (Ty1-Ty5), four are are cell-type regulated (Ty1, Ty2, Ty3 and Ty5); that is, their transcription is repressed in diploid cells. For Ty1-Ty3, the repression in diploid cells is due to the presence of an a1/α2 repressor complex binding site within their transcription regulatory sequences. It's not clear if this is also the reason for diploid repression of Ty5 although the same site is present in the Ty5 LTR (Ke et al., 1997). Ty3 and Ty5 mobility is activated when haploid cells are exposed to mating pheromones (Ke et al., 1997; Kinsey and Sandmeyer, 1995). Several pheromone responsive elements (PRE) are present in their 5' LTR, explaining their activation by the mating pheromones. In contrast, the PRE sites are missing in Ty1, which corresponds to its inability to be activated by mating pheromones. The transposition of Ty1, however, can be activated when it is exposed to external signals that induce invasive/filamentous differentiation. A FRE (filamentous responsive element) binding site is located within the TY1A coding sequence (Baur et al., 1997). Ty1 can also be activated by DNA lesions. The most compelling evidence comes from the activation of Ty1 retrotransposition by telomere erosion (Scholes et al., 2003). Ty1 retrotransposition increases in parallel with telomere erosion and then partially declines when survivors emerge. Exposure of yeast cells to a variety of DNA damaging agents (UV light, γ irradiation and chemicals) also increases Ty1 mobility (Staleva Staleva and Venkov, 2001). Other environmental challenges might activate Ty1 transposition via GCN4, the master regulator of gene expression under a large spectrum of stress conditions (Natarajan et al., 2001). Five GCN4 binding sites were found in the 5' UTR of the Ty1 and Ty2 families, and GCN4 is necessary for their transcription under the condition of amino acid starvation (Morillon et al., 2002).

THE INTEGRATION SPECIFICITY OF TES

One aspect of TE biology, which interests lots of researchers, is where does the element integrate within the host genome and why? Targeted integration into certain chromatin regions likely serves to facilitate element propagation and thereby optimizes the element-host relationship. The wide variety of patterns of target site selection suggests that many strategies successfully promote these objectives (Craig, 1997). Extensive work has been done on retroviruses, especially the human immunodeficiency virus (HIV), and on some retrotransposons, and this work has shed light on several possible mechanisms used by different TEs.

Early studies suggest that the accessibility of the target DNA might affect integration site selection. 'Open chromatin,' such as transcribed regions, can facilitate access of retroviral integration complexes, and therefore become the preferential integration sites. Two groups reported that Moloney murine leukemia virus (MLV) retroviruses preferentially integrate close to DNase I-hypersensitive sites *in vivo* (Rohdewohld et al., 1987; Vijaya et al., 1986). However, this preference seems to virus-specific, since similar sites are not favored for HIV integration (Ciuffi and Bushman, 2006).

Some studies suggest that chromosomal regions containing sequences that are preferentially recognized and cleaved by the element-encoded endonuclease (integrase) may be the specific target sites for those elements. It seems to be true for some of the non-LTR retrotransposons, such as the R2 element in *Bombyx mori* and L1 elements in human (Cost and Boeke, 1998; Feng et al., 1996; Luan et al., 1993). The expressed R2 ORF encodes both sequence-specific endonuclease and reverse transcriptase activities and is able to recognize a specific sequence in 28S genes (Luan et al., 1993). The endonuclease domain (EN) in L1 ORF2 preferentially cleaves sequences resembling L1 *in vivo* target sequences (Cost and Boeke, 1998; Feng et al., 1996).

Another factor which might affect integration specificity is the point during the cell cycle when integration takes places (Bushman et al., 2005). Recently, Barr et al. analyzed more than 750 unique HIV integration sites in mature macrophage, which do not divide (Barr

et al., 2006). They found that integration is generally favored in active transcription units, as seen in dividing PBMC or T cell lines, suggesting that the timing during the cell cycle when HIV integrates does not affect integration site selection. They also found that two different envelopes (VSV-G and CCR5-tropic (R5) HIV Bal envelopes) used for infection showed no significant differences in the distribution of integration sites. Therefore, the envelope used for infection and the route of entry do not strongly affect integration site selection. These results are consistent with a previous study of integration in nondividing cells, in which HIV vector integration in growth-arrested IMR-90 lung fibroblasts was also found to be favored in transcription units (Ciuffi et al., 2006).

Another model for target site selection is the tethering model, which invokes direct contacts between retroviral integration complexes and cellular proteins bound specifically at favored target sites (Ciuffi and Bushman, 2006). In principle, any component of the integration complex or its associated proteins could serve as the docking point for a tethering factor that directs integration. This model is supported by lots of data from the retrovirus HIV (reviewed by (Ciuffi and Bushman, 2006)) and the Ty retrotransposons in yeast (Bushman, 2003; Kirchner et al., 1995; Sandmeyer, 2003; Zhu et al., 2003). Genome-wide surveys of integration sites for HIV and MLV in humans have been performed (Schroder et al., 2002; Wu et al., 2003). Both reports found that transcription units, especially active genes, are favored integration sites for HIV and MLV in infected target cells. A similar pattern was revealed in a study of SIV integration in macaques (Hematti et al., 2004), suggesting that targeted integration into active genes might be a common feature of lentiviruses, possibly due to a similar targeting mechanism. Recently, the cellular lens epithelium-derived growth factor (PSIP1/LEDGF/p75) protein, which binds tightly to HIV integrase (Cherepanov et al., 2003; Turlure et al., 2004), has been found to influence the placement of HIV integration sites *in vivo*, providing the first example of a cellular protein controlling the location of HIV integration in human cells (Ciuffi et al., 2005). However, knockdown of LEDGF/p75 did not fully eliminate favored integration in transcription units, leaving open the possibility that additional cellular factors may be involved in targeting HIV DNA integration (Ciuffi et al., 2005).

Targeting is particularly pronounced for the five retrotransposon families of budding yeast (recently reviewed by (Lesage and Todeschini, 2005)). Ty3 integrates specifically into the transcription initiation sites of genes transcribed by RNA pol III, including tRNA, 5S and U6 genes. The integration sites are restricted within a few nucleotides of the transcription start site (Chalker and Sandmeyer, 1992). Targeted integration of Ty3 is likely mediated by interactions between the integration complex and pol III transcription factors, because it has been shown that TFIIIB and TFIIIC are required for position-specific integration (Kirchner et al., 1995). Ty1 shares a similar integration pattern with Ty3 and also targets to pol III transcribed genes. However, Ty1 target sites are not as tight as Ty3's; rather, they integrate into a window spanning about 600bp upstream of the transcription initiation sites (Boeke and Devine, 1998; Devine and Boeke, 1996). Although both Ty1 and Ty3 integrate into a similar region, their targeting mechanism may not be the same, as suggested by their significantly different integration pattern. It is still possible that Ty1 targeting depends on the binding of some pol III related factors. There is little known about Ty2 and Ty4 targeting except that their insertions are also associated with pol III genes and share a pattern similar to that of Ty1 (Kim et al., 1998). Ty5, on the other hand, has a totally different integration pattern from all the other Tys, and work with Ty5 provides the best support for the tethering model. Detailed discussion of Ty5 is presented in the following section.

THE TY5 RETROTRANSPOSON IN YEAST

The full length Ty5 retrotransposon is about 5 kb, flanked by two 251 bp LTRs. It encodes a single ORF, which contains Gag and Pol (Figure 1) (Gao et al., 2002). After translation, the polyprotein is processed by the element-encoded protease, generating the mature Gag proteins, IN and RT/RH. There are two forms of Gag proteins with sizes of 27 kD and 37 kD. The IN and RT/RH are about 80 kD and 59 kD, respectively. All the Ty5 proteins are insoluble (Irwin and Voytas, 2001).

Active Ty5 elements have been lost from *S. cerevisiae*, leaving only one non-functional copy of Ty5 that resides at the telomere of chromosome III and six solo LTRs (Kim et al., 1998; Voytas and Boeke, 1992). With only seven elements, the Ty5 elements are the least abundant *S. cerevisiae* transposon family. Four of the solo LTRs reside near the

telomeres and the other two are near *HMR* (Kim et al., 1998; Zou et al., 1995). The Ty5 elements have also been identified in the closely related species *S. paradoxus*. Numerous Ty5 insertions were found in *S. paradoxus*, and most are also located in silent chromatin (Zou et al., 1995). An active copy of Ty5 was retrieved from *S. paradoxus* and modified for regulated expression in S. cerevisiae, making it possible to study its *de novo* integration in *this model yeast.* (Zou et al., 1996a). However, the transposition efficiency is very low (about 10^{-6} to 10^{-4} per element per cell). Nonetheless, 14 of 15 newly transposed Ty5 elements were found to map to telomeric regions on 10 different chromosomes. Nine of these insertions are within 0.8 kb and three are within 1.5 kb of the autonomously replicating consensus sequence in the subtelomeric X repeat. This suggests that the X repeat plays an important role in directing Ty5 integration (Zou et al., 1996b). Ty5 can also target to the other heterochromatic regions, namely the silent mating loci, and integration here accounts for ~2% of the total transposition events. Most insertions occur within 800 bp on either side of the autonomously replicating consensus sequence within *HMR-E.* Silencer mutations which alleviate transcriptional repression also abolish integration near HMR-E, suggesting the protein complexes assembled at the silencer may direct integration (Zou and Voytas, 1997).

Further analysis demonstrated that loss of two components of silent chromatin, Sir3p or Sir4p, significantly decreases Ty5 targeting to the *HM* loci and largely randomizes chromosomal integration patterns. In strains with the *sir4-42 allele*, in which the Sir complex relocates to rDNA, about 26% of Ty5 insertions occur within the rDNA. This contrasts with wild type strains in which 3% of the insertions occur in the rDNA (Zhu et al., 1999). It should be pointed out that deletion of SIR2 does not randomize Ty5 integration, although silencing is reduced in *sir2* mutants to a level observed in *sir3* or *sir4* strains. This observation suggests that silencing *per se* is not required for targeting. Instead, the proteins that maintain silencing, such as Sir3p or Sir4p, are likely the players that direct Ty5 integration. This also explains the change of targeting observed in *sir4-42* strains.

Mutagenesis experiments mapped the Ty5-encoded targeting determinant to the C-terminus of Ty5 IN (INC). A single amino acid substitution in INC abolishes Ty5 integration specificity (Gai and Voytas, 1998). Further analysis defined the targeting domain (TD) in

INC. TD spans six amino acids, four of which (in bold) are required for targeted integration (**LDSSPP**) (Xie et al., 2001). Direct interaction between Ty5 INC and Sir4C was observed in the yeast two-hybrid assay and *in vitro* binding assays (Xie et al., 2001). Furthermore, Ty5 integration hotspots can be created when Sir4p is tethered to ectopic DNA sites, suggesting that the interaction between Sir4C and IN may be the sole determinant of Ty5 target specificity (Zhu et al., 2003). In addition, the target specificity of Ty5 can be altered by swapping TD with exogenous protein binding motifs (Zhu et al., 2003). This collective work increases our understanding of the targeting mechanisms of Ty5 retrotransposons and provides concrete support for the tethered integration hypothesis.

DISSERTATION ORGANIZATION

Chapter II was published in *Proc Natl Acad Sci USA. The paper was* highlighted on the cover and reviewed by both *Proc Natl Acad Sci U S A.* 100:5586-8 (Integration by Design, Sandmeyer, S.) and *Cell* 115:135-138 (Targeting Survival: Integration Site Selection by Retroviruses and LTR-Retrotransposons. Bushman, F.D.). In this paper, we show that new Ty5 integration hot spots are created when Sir4p is tethered to ectopic DNA sites. Targeting to sites of tethered Sir4p is abrogated by single amino acid substitutions in either IN or Sir4p that prevent their interaction. Ty5 target specificity can be altered by replacing the IN-targeting domain with other peptide motifs that interact with known protein partners. Integration occurs at high efficiency and in close proximity to DNA sites where the protein partners are tethered. These findings define a mechanism by which retrotransposons shape their host genomes and suggest ways in which retroviral integration can be controlled. The work in this chapter was carried out collaboratively with another former graduate student, Yunxia Zhu. Yunxia established the assay system to test for integration to sites of tethered proteins, and I was able to successfully modify the Ty5 element and redirect its integration. Chapter III is a manuscript prepared for submission to *Molecular Cell.* The paper presents evidence that the Ty5 targeting domain is post-translationally modified *in vivo.* The modification (phosphorylation) is required for targeted Ty5 integration, as measured by our targeting assay. Using surface plasmon resonance (SPR) spectroscopy, we found that phosphorylation of TD is required for its productive interaction with Sir4C *in vitro.* This provides direct evidence for the requirement for post-translational modification in Ty5

targeting. A total of 109 kinase deletion strains were screened for kinases affecting Ty5 target specificity. Among eight candidates, Dun1p was found to be the most likely to modify the Ty5 targeting domain. This project was initiated by the former graduate student, Weiwu Xie, who carried out the screening of the kinase deletion strains. *In vitro* tests of the interaction between TD and Sir4C was done by another graduate student, Troy Brady. I carried out the rest of work. Chapter IV is a manuscript prepared for submission to *Genetics*. In this manuscript, I surveyed the effects on silencing when various kinases are knocked out from the yeast genome. Silencing at three loci – the telomeres, silent mating loci and rDNA – is evaluated in each kinase deletion background. I carried out all the work in this chapter. Chapter V is a project I initiated to examine the relationship between phosphorylation and Ty5 transposition and/or targeting. A complete set of yeast GST-kinase fusion proteins were expressed and purified individually and subjected to *in vitro* kinase assays with various substrates purified from *E. coli*. We found 16 kinases that are able to phosphorylate the wild type integrase C-terminus to different extents. Most of these kinases affected Ty5 transposition and/or targeting when they were knocked out. We are currently mapping the sites of phosphorylation for each kinase. This project will be continued by a new graduate student, Jiquan Gao and we mutually agree that we will share the first authorship when the paper gets published. Chapter VI provides a summary of my research and a general discussion.

REFERENCES

Aravin, A. A., Klenov, M. S., Vagin, V. V., Bantignies, F., Cavalli, G., and Gvozdev, V. A. (2004). Dissection of a natural RNA silencing process in the Drosophila melanogaster germ line. Mol Cell Biol *24*, 6742-6750.

Barr, S. D., Ciuffi, A., Leipzig, J., Shinn, P., Ecker, J. R., and Bushman, F. D. (2006). HIV Integration Site Selection: Targeting in Macrophages and the Effects of Different Routes of Viral Entry. Mol Ther.

Baur, M., Esch, R. K., and Errede, B. (1997). Cooperative binding interactions required for function of the Ty1 sterile responsive element. Mol Cell Biol *17*, 4330-4337.

Bejerano, G., Lowe, C. B., Ahituv, N., King, B., Siepel, A., Salama, S. R., Rubin, E. M., Kent, W. J., and Haussler, D. (2006). A distal enhancer and an ultraconserved exon are derived from a novel retroposon. Nature *441*, 87-90.

Bennetzen, J. L. (2005). Transposable elements, gene creation and genome rearrangement in flowering plants. Curr Opin Genet Dev *15*, 621-627.

Blackburn, E. H. (1992). Telomerases. Annu Rev Biochem *61*, 113-129.

Boeke, J. D., and Devine, S. E. (1998). Yeast retrotransposons: finding a nice quiet neighborhood. Cell *93*, 1087-1089.

Britten, R. (2006). Transposable elements have contributed to thousands of human proteins. Proc Natl Acad Sci U S A *103*, 1798-1803.

Buchon, N., and Vaury, C. (2006). RNAi: a defensive RNA-silencing against viruses and transposable elements. Heredity *96*, 195-202.

Bushman, F., Lewinski, M., Ciuffi, A., Barr, S., Leipzig, J., Hannenhalli, S., and Hoffmann, C. (2005). Genome-wide analysis of retroviral DNA integration. Nat Rev Microbiol *3*, 848-858.

Bushman, F. D. (2003). Targeting survival: integration site selection by retroviruses and LTR-retrotransposons. Cell *115*, 135-138.

Chalker, D. L., and Sandmeyer, S. B. (1992). Ty3 integrates within the region of RNA polymerase III transcription initiation. Genes Dev *6*, 117-128.

Cherepanov, P., Maertens, G., Proost, P., Devreese, B., Van Beeumen, J., Engelborghs, Y., De Clercq, E., and Debyser, Z. (2003). HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. J Biol Chem *278*, 372-381.

Chicas, A., Cogoni, C., and Macino, G. (2004). RNAi-dependent and RNAi-independent mechanisms contribute to the silencing of RIPed sequences in Neurospora crassa. Nucleic Acids Res *32*, 4237-4243.

Ciuffi, A., and Bushman, F. D. (2006). Retroviral DNA integration: HIV and the role of LEDGF/p75. Trends Genet.

Ciuffi, A., Llano, M., Poeschla, E., Hoffmann, C., Leipzig, J., Shinn, P., Ecker, J. R., and Bushman, F. (2005). A role for LEDGF/p75 in targeting HIV DNA integration. Nat Med *11*, 1287-1289.

Ciuffi, A., Mitchell, R. S., Hoffmann, C., Leipzig, J., Shinn, P., Ecker, J. R., and Bushman, F. D. (2006). Integration site selection by HIV-based vectors in dividing and growth-arrested IMR-90 lung fibroblasts. Mol Ther *13*, 366-373.

Cost, G. J., and Boeke, J. D. (1998). Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. Biochemistry *37*, 18081-18093.

Craig, N. L. (1997). Target site selection in transposition. Annu Rev Biochem *66*, 437-474.

Devine, S. E., and Boeke, J. D. (1996). Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. Genes Dev *10*, 620-633.

Engels, W. R., Johnson-Schlitz, D. M., Eggleston, W. B., and Sved, J. (1990). High-frequency P element loss in Drosophila is homolog dependent. Cell *62*, 515-525.

Feng, Q., Moran, J. V., Kazazian, H. H., Jr., and Boeke, J. D. (1996). Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell *87*, 905-916.

Freitag, M., Hickey, P. C., Khlafallah, T. K., Read, N. D., and Selker, E. U. (2004). HP1 is essential for DNA methylation in neurospora. Mol Cell *13*, 427-434.

Gai, X., and Voytas, D. F. (1998). A single amino acid change in the yeast retrotransposon Ty5 abolishes targeting to silent chromatin. Mol Cell *1*, 1051-1055.

Gao, X., Rowley, D. J., Gai, X., and Voytas, D. F. (2002). Ty5 gag mutations increase retrotransposition and suggest a role for hydrogen bonding in the function of the nucleocapsid zinc finger. J Virol *76*, 3240-3247.

Gendrel, A. V., Lippman, Z., Yordan, C., Colot, V., and Martienssen, R. A. (2002). Dependence of heterochromatic histone H3 methylation patterns on the Arabidopsis gene DDM1. Science *297*, 1871-1873.

Grandbastien, M. A., Audeon, C., Bonnivard, E., Casacuberta, J. M., Chalhoub, B., Costa, A. P., Le, Q. H., Melayah, D., Petit, M., Poncet, C., *et al.* (2005). Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae. Cytogenet Genome Res *110*, 229-241.

Gray, Y. H. (2000). It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. Trends Genet *16*, 461-468.

Han, J. S., Szak, S. T., and Boeke, J. D. (2004). Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. Nature *429*, 268-274.

Harley, C. B., Futcher, A. B., and Greider, C. W. (1990). Telomeres shorten during ageing of human fibroblasts. Nature *345*, 458-460.

Hematti, P., Hong, B. K., Ferguson, C., Adler, R., Hanawa, H., Sellers, S., Holt, I. E., Eckfeldt, C. E., Sharma, Y., Schmidt, M., *et al.* (2004). Distinct genomic integration of MLV and SIV vectors in primate hematopoietic stem and progenitor cells. PLoS Biol *2*, e423.

Initiative, T. A. G. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature *408*, 796-815.

Irwin, B., Aye, M., Baldi, P., Beliakova-Bethell, N., Cheng, H., Dou, Y., Liou, W., and Sandmeyer, S. (2005). Retroviruses and yeast retrotransposons use overlapping sets of host genes. Genome Res *15*, 641-654.

Irwin, P. A., and Voytas, D. F. (2001). Expression and processing of proteins encoded by the Saccharomyces retrotransposon Ty5. J Virol *75*, 1790-1797.

Jiang, J., Birchler, J. A., Parrott, W. A., and Dawe, R. K. (2003). A molecular view of plant centromeres. Trends Plant Sci *8*, 570-575.

Kalendar, R., Vicient, C. M., Peleg, O., Anamthawat-Jonsson, K., Bolshoy, A., and Schulman, A. H. (2004). Large retrotransposon derivatives: abundant, conserved but nonautonomous retroelements of barley and related genomes. Genetics *166*, 1437-1450.

Kapitonov, V. V., and Jurka, J. (1999). Molecular paleontology of transposable elements from Arabidopsis thaliana. Genetica *107*, 27-37.

Kato, M., Miura, A., Bender, J., Jacobsen, S. E., and Kakutani, T. (2003). Role of CG and non-CG methylation in immobilization of transposons in Arabidopsis. Curr Biol *13*, 421-426.

Kato, M., Takashima, K., and Kakutani, T. (2004). Epigenetic control of CACTA transposon mobility in Arabidopsis thaliana. Genetics *168*, 961-969.

Ke, N., Irwin, P. A., and Voytas, D. F. (1997). The pheromone response pathway activates transcription of Ty5 retrotransposons located within silent chromatin of Saccharomyces cerevisiae. Embo J *16*, 6272-6280.

Ketting, R. F., Haverkamp, T. H., van Luenen, H. G., and Plasterk, R. H. (1999). Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell *99*, 133-141.

Kim, J. M., Vanguri, S., Boeke, J. D., Gabriel, A., and Voytas, D. F. (1998). Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete Saccharomyces cerevisiae genome sequence. Genome Res *8*, 464-478.

Kinsey, P. T., and Sandmeyer, S. B. (1995). Ty3 transposes in mating populations of yeast: a novel transposition assay for Ty3. Genetics *139*, 81-94.

Kirchner, J., Connolly, C. M., and Sandmeyer, S. B. (1995). Requirement of RNA polymerase III transcription factors for in vitro position-specific integration of a retroviruslike element. Science *267*, 1488-1491.

Kupiec, M., and Steinlauf, R. (1997). Damage-induced ectopic recombination in the yeast Saccharomyces cerevisiae. Mutat Res *384*, 33-44.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Langdon, T., Seago, C., Mende, M., Leggett, M., Thomas, H., Forster, J. W., Jones, R. N., and Jenkins, G. (2000). Retrotransposon evolution in diverse plant genomes. Genetics *156*, 313-325.

Lemoine, F. J., Degtyareva, N. P., Lobachev, K., and Petes, T. D. (2005). Chromosomal translocations in yeast induced by low levels of DNA polymerase a model for chromosome fragile sites. Cell *120*, 587-598.

Lesage, P., and Todeschini, A. L. (2005). Happy together: the life and times of Ty retrotransposons and their hosts. Cytogenet Genome Res *110*, 70-90.

Li, W., Zhang, P., Fellers, J. P., Friebe, B., and Gill, B. S. (2004). Sequence composition, organization, and evolution of the core Triticeae genome. Plant J *40*, 500-511.

Lohe, A. R., Moriyama, E. N., Lidholm, D. A., and Hartl, D. L. (1995). Horizontal transmission, vertical inactivation, and stochastic loss of mariner-like transposable elements. Mol Biol Evol *12*, 62-72.

Luan, D. D., Korman, M. H., Jakubczak, J. L., and Eickbush, T. H. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell *72*, 595-605.

Maxwell, P. H., Coombes, C., Kenny, A. E., Lawler, J. F., Boeke, J. D., and Curcio, M. J. (2004). Ty1 mobilizes subtelomeric Y' elements in telomerase-negative Saccharomyces cerevisiae survivors. Mol Cell Biol *24*, 9887-9898.

Mc, C. B. (1950). The origin and behavior of mutable loci in maize. Proc Natl Acad Sci U S A *36*, 344-355.

Messing, J., Bharti, A. K., Karlowski, W. M., Gundlach, H., Kim, H. R., Yu, Y., Wei, F., Fuks, G., Soderlund, C. A., Mayer, K. F., and Wing, R. A. (2004). Sequence composition and genome organization of maize. Proc Natl Acad Sci U S A *101*, 14349-14354.

Meyers, B. C., Tingey, S. V., and Morgante, M. (2001). Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. Genome Res *11*, 1660-1676.

Miura, A., Yonebayashi, S., Watanabe, K., Toyama, T., Shimada, H., and Kakutani, T. (2001). Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. Nature *411*, 212-214.

Morillon, A., Benard, L., Springer, M., and Lesage, P. (2002). Differential effects of chromatin and Gcn4 on the 50-fold range of expression among individual yeast Ty1 retrotransposons. Mol Cell Biol *22*, 2078-2088.

Nagaki, K., Cheng, Z., Ouyang, S., Talbert, P. B., Kim, M., Jones, K. M., Henikoff, S., Buell, C. R., and Jiang, J. (2004). Sequencing of a rice centromere uncovers active genes. Nat Genet *36*, 138-145.

Nagaki, K., Song, J., Stupar, R. M., Parokonny, A. S., Yuan, Q., Ouyang, S., Liu, J., Hsiao, J., Jones, K. M., Dawe, R. K., *et al.* (2003). Molecular and cytological analyses of large tracks of centromeric DNA reveal the structure and evolutionary dynamics of maize centromeres. Genetics *163*, 759-770.

Nakagawa, H., Lee, J. K., Hurwitz, J., Allshire, R. C., Nakayama, J., Grewal, S. I., Tanaka, K., and Murakami, Y. (2002). Fission yeast CENP-B homologs nucleate centromeric heterochromatin by promoting heterochromatin-specific histone tail modifications. Genes Dev *16*, 1766-1778.

Natarajan, K., Meyer, M. R., Jackson, B. M., Slade, D., Roberts, C., Hinnebusch, A. G., and Marton, M. J. (2001). Transcriptional profiling shows that Gcn4p is a master regulator of gene expression during amino acid starvation in yeast. Mol Cell Biol *21*, 4347-4368.

Nekrutenko, A., and Li, W. H. (2001). Transposable elements are found in a large number of human protein-coding genes. Trends Genet *17*, 619-621.

Pardue, M. L., Rashkova, S., Casacuberta, E., DeBaryshe, P. G., George, J. A., and Traverse, K. L. (2005). Two retrotransposons maintain telomeres in Drosophila. Chromosome Res *13*, 443-453.

Peterson-Burch, B. D., Wright, D. A., Laten, H. M., and Voytas, D. F. (2000). Retroviruses in plants? Trends Genet *16*, 151-152.

Rohdewohld, H., Weiher, H., Reik, W., Jaenisch, R., and Breindl, M. (1987). Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. J Virol *61*, 336-343.

Sabot, F., Guyot, R., Wicker, T., Chantret, N., Laubin, B., Chalhoub, B., Leroy, P., Sourdille, P., and Bernard, M. (2005). Updating of transposable element annotations from large wheat genomic sequences reveals diverse activities and gene associations. Mol Genet Genomics *274*, 119-130.

Sandmeyer, S. (2003). Integration by design. Proc Natl Acad Sci U S A *100*, 5586-5588.

SanMiguel, P., Tikhonov, A., Jin, Y. K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P. S., Edwards, K. J., Lee, M., Avramova, Z., and Bennetzen, J. L. (1996). Nested retrotransposons in the intergenic regions of the maize genome. Science *274*, 765-768.

Schatz, D. G., and Spanopoulou, E. (2005). Biochemistry of V(D)J recombination. Curr Top Microbiol Immunol *290*, 49-85.

Scholes, D. T., Kenny, A. E., Gamache, E. R., Mou, Z., and Curcio, M. J. (2003). Activation of a LTR-retrotransposon by telomere erosion. Proc Natl Acad Sci U S A *100*, 15736-15741.

Schroder, A. R., Shinn, P., Chen, H., Berry, C., Ecker, J. R., and Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. Cell *110*, 521-529.

Schulman, A. H., and Kalendar, R. (2005). A movable feast: diverse retrotransposons and their contribution to barley genome dynamics. Cytogenet Genome Res *110*, 598-605.

Slotkin, R. K., Freeling, M., and Lisch, D. (2005). Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. Nat Genet *37*, 641-644.

Smit, A. F. (1999). Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr Opin Genet Dev *9*, 657-663.

Smit, A. F., and Riggs, A. D. (1996). Tiggers and DNA transposon fossils in the human genome. Proc Natl Acad Sci U S A *93*, 1443-1448.

Spicuglia, S., Franchini, D. M., and Ferrier, P. (2006). Regulation of V(D)J recombination. Curr Opin Immunol *18*, 158-163.

Staleva Staleva, L., and Venkov, P. (2001). Activation of Ty transposition by mutagens. Mutat Res *474*, 93-103.

Takahashi, H., and Fujiwara, H. (1999). Transcription analysis of the telomeric repeat-specific retrotransposons TRAS1 and SART1 of the silkworm Bombyx mori. Nucleic Acids Res *27*, 2015-2021.

Tijsterman, M., Ketting, R. F., Okihara, K. L., Sijen, T., and Plasterk, R. H. (2002). RNA helicase MUT-14-dependent gene silencing triggered in C. elegans by short antisense RNAs. Science *295*, 694-697.

Turlure, F., Devroe, E., Silver, P. A., and Engelman, A. (2004). Human cell proteins and human immunodeficiency virus DNA integration. Front Biosci *9*, 3187-3208.

Vijaya, S., Steffen, D. L., and Robinson, H. L. (1986). Acceptor sites for retroviral integrations map near DNase I-hypersensitive sites in chromatin. J Virol *60*, 683-692.

Voytas, D. F., and Boeke, J. D. (1992). Yeast retrotransposon revealed. Nature *358*, 717.

Wheelan, S. J., Aizawa, Y., Han, J. S., and Boeke, J. D. (2005). Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution. Genome Res *15*, 1073-1078.

Wright, D. A., and Voytas, D. F. (1998). Potential retroviruses in plants: Tat1 is related to a group of Arabidopsis thaliana Ty3/gypsy retrotransposons that encode envelope-like proteins. Genetics *149*, 703-715.

Wu, X., Li, Y., Crise, B., and Burgess, S. M. (2003). Transcription start regions in the human genome are favored targets for MLV integration. Science *300*, 1749-1751.

Wu-Scharf, D., Jeong, B., Zhang, C., and Cerutti, H. (2000). Transgene and transposon silencing in Chlamydomonas reinhardtii by a DEAH-box RNA helicase. Science *290*, 1159-1162.

Xie, W., Gai, X., Zhu, Y., Zappulla, D. C., Sternglanz, R., and Voytas, D. F. (2001). Targeting of the yeast Ty5 retrotransposon to silent chromatin is mediated by interactions between integrase and Sir4p. Mol Cell Biol *21*, 6606-6614.

Yoder, J. A., Walsh, C. P., and Bestor, T. H. (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends Genet *13*, 335-340.

Zhang, J., and Peterson, T. (1999). Genome rearrangements by nonlinear transposons in maize. Genetics *153*, 1403-1410.

Zhang, J., and Peterson, T. (2005). A segmental deletion series generated by sister-chromatid transposition of Ac transposable elements in maize. Genetics *171*, 333-344.

Zhu, Y., Dai, J., Fuerst, P. G., and Voytas, D. F. (2003). Controlling integration specificity of a yeast retrotransposon. Proc Natl Acad Sci U S A *100*, 5891-5895.

Zhu, Y., Zou, S., Wright, D. A., and Voytas, D. F. (1999). Tagging chromatin with retrotransposons: target specificity of the Saccharomyces Ty5 retrotransposon changes with the chromosomal localization of Sir3p and Sir4p. Genes Dev *13*, 2738-2749.

Zilberman, D., Cao, X., and Jacobsen, S. E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science *299*, 716-719.

Zou, S., Ke, N., Kim, J. M., and Voytas, D. F. (1996a). The Saccharomyces retrotransposon Ty5 integrates preferentially into regions of silent chromatin at the telomeres and mating loci. Genes Dev *10*, 634-645.

Zou, S., Kim, J. M., and Voytas, D. F. (1996b). The Saccharomyces retrotransposon Ty5 influences the organization of chromosome ends. Nucleic Acids Res *24*, 4825-4831.

Zou, S., and Voytas, D. F. (1997). Silent chromatin determines target preference of the Saccharomyces retrotransposon Ty5. Proc Natl Acad Sci U S A *94*, 7412-7416.

Zou, S., Wright, D. A., and Voytas, D. F. (1995). The Saccharomyces Ty5 retrotransposon family is associated with origins of DNA replication at the telomeres and the silent mating locus HMR. Proc Natl Acad Sci U S A *92*, 920-924.

CHAPTER 2. CONTROLLING INTEGRATION SPECIFICITY OF A YEAST RETROTRANSPOSON

A paper published in *Proc Natl Acad Sci U S A*[1]

Yunxia Zhu[2], Junbiao Dai[3], Peter G. Fuerst[4] and Daniel F. Voytas[5]

ABSTRACT

Retrotransposons and retroviruses integrate non-randomly into eukaryotic genomes. For the yeast retrotransposon Ty5, integration preferentially occurs within domains of heterochromatin. Targeting to these locations is determined by interactions between an amino acid sequence motif at the C-terminus of Ty5 integrase (the targeting domain) and the heterochromatin protein Sir4p. Here we show that new Ty5 integration hotspots are created when Sir4p is tethered to ectopic DNA sites. Targeting to sites of tethered Sir4p is abrogated by single amino acid substitutions in either integrase or Sir4p that prevent their interaction. Ty5 target specificity can be altered by replacing the integrase targeting domain with other peptide motifs that interact with known protein partners. Integration occurs at high efficiency and in close proximity to DNA sites where the protein partners are tethered. These findings define a mechanism by which retrotransposons shape their host genomes and suggest ways in which retroviral integration can be controlled.

INTRODUCTION

cDNA synthesized by retroviruses and retrotransposons is integrated into host genomes by the retroelement-encoded integrase protein. Integration is essential for retroviral proliferation and has significantly shaped eukaryotic genome organization. For example, endogenous retroviruses and retrotransposons constitute over half the genomes of human and

---

[1] Reprinted with permission from *PNAS*, 2003, 100:5891-5

[2] First author who established the assay system to test for integration to sites of tethered proteins

[3] Second and thesis author who modified the Ty5 element and redirected its integration

[4] Third author who contributed some of the two-hybrid assay

[5] Professor and corresponding author, Department of Genetics, Development and Cell Biology, Iowa State University, Ames, IA 50011

some plant species such as maize (1, 2). Retroelement insertions are not randomly distributed within genomes, and they are often enriched in heterochromatin or other gene-poor regions. This distribution may be due, in part, to selection against integration within gene-rich euchromatin. Alternatively, retroelements may actively select targets such as heterochromatin where insertion will not compromise host fitness (3, 4).

Integration site choice is clearly the rule for the well-studied Ty elements of *Saccharomyces cerevisiae*. Ty1 and Ty3 both integrate preferentially upstream of genes transcribed by RNA polymerase III (e.g. tRNA genes), which are harmless sites because they are gene-poor and integration does not disrupt Pol III transcription (5, 6). The precision of integration, however, differs for these two yeast retroelements. Ty3 typically integrates within 1-2 bases of Pol III transcription start sites, whereas Ty1 inserts within a 750 base pair window upstream of target genes. Pol III complexes are required for both Ty1 and Ty3 integration specificity, and for Ty3, TFIIIB is the major determinant of target choice (7-10). This suggests that target sites are selected through interactions between Ty1 and Ty3 integration complexes and the Pol III transcription machinery.

In contrast to Ty1 and Ty3 and like many other eukaryotic retrotransposons, Ty5 insertions are found within heterochromatic regions of the yeast genome (11). This distribution is due to target site selection, since 95% of *de novo* Ty5 transposition events occur within heterochromatin found at yeast telomeres and the silent mating loci (*HML* and *HMR*) (12, 13). Several lines of evidence suggest that the heterochromatin protein Sir4p is the major determinant of Ty5 integration specificity (14, 15). A six amino acid motif at the C-terminus of Ty5 integrase (the targeting domain) interacts with Sir4p (15), and mutations in the targeting domain that abrogate Sir4p interactions randomize Ty5 integration patterns (15, 16). Similarly, Ty5 integrates randomly in strains that lack Sir4p (14).

A recent study revealed that HIV integration occurs preferentially at sites of active transcription (17). Furthermore, HIV integrase interacts with Ini1, a homolog of the yeast transcriptional activator Snf5p (18). These observations, coupled with the data for the Ty retrotransposons, suggest a general model wherein interactions between integrase and DNA-

bound proteins mediate retroelement target choice. By further defining the determinants of Ty5 integration specificity and by engineering Ty5 elements with altered target site preference, we demonstrate that this model describes the mechanism by which Ty5 selects integration sites.

METHODS

Plasmid constructs.

DNA fragments encoding various regions of the Sir4p C-terminus were amplified by the polymerase chain reaction (PCR) (19) from plasmid pRS316-*SIR4* (gift of Jasper Rine, University of California, Berkeley). The amplification products were cloned into the *Eco*RI/*Bam*HI sites of the LexA-expressing vector pBTM116 (20) or a derivative with a *LEU2* marker gene (pYZ275). PCR mutagenesis was used to substitute an alanine at residues 971 to 976 of Sir4p (19). Additional LexA fusion constructs were generated by PCR-amplifying the FHA1 domain of Rad53p (amino acids 14-154) from yeast genomic DNA and the coding region for Npw38 from a human cDNA clone (ATCC5806979). PCR products were cloned into the *Eco*RI and *Bgl*II sites of pYZ275. The C-terminus of Ty5 integrase (amino acids 934-1131) was amplified by PCR from pNK254 (21) or a variant with the S1094L targeting domain mutation (16). The amplification products were cloned into the *Xma*I/*Pst*I sites of pGAD-C1 (22) to generate fusions with the Gal4p activation domain.

The donor plasmids carry either a wild type Ty5 element or one with the S1094L mutation. Both Ty5 elements are under transcriptional control of the *GAL1-10* promoter and carry a *his3AI* selectable marker gene (23). The six amino acid targeting domain of Ty5 (LDSSPP) was replaced with a motif of Rad9p (SLEVTEADATFVQ) (24) to generate Ty5-Rad9p or a motif of NpwBP (PRLLPPFPPPGR) (25) to generate Ty5-NpwBP. These modifications were made using a two-step PCR replacement method (19). The target plasmid was generated by inserting 3-kb and 4-kb DNA fragments from Arabidopsis into the *Eco*RI site and *Sac*I sites of pRS424 (26), which was modified to carry a Chl^r gene. One to four copies of overlapping, double LexA operators (20) were inserted into the *Bam*HI site.

Tethered integration and two-hybrid assays.

YPH499 or its *sir* derivatives were transformed with a Ty5 donor plasmid, a target plasmid, and a plasmid expressing a LexA fusion protein. Yeast cells were grown as patches on media lacking tryptophan, leucine and uracil (SC-T-L-U) and incubated at 30°C for two days. The cells were then replicated onto the same selective media supplemented with 2% galactose and incubated at room temperature for two days. Finally, the cells were replica plated onto media that lacked histidine (SC-T-L-U-H) and incubated at 30°C for three days. Cells were scraped from the plates, washed twice with water, and total DNA was prepared (19). The DNA was used to transform competent *E. coli* cells with a *hisB* mutation (strain eDW335, Wright and Voytas unpublished). After transformation, the *E. coli* cells were incubated in rich media at 37°C for three hours and washed twice with water to remove the residual histidine. One-tenth of the cells were plated onto rich media with 20 μg/μl chloramphenicol and nine-tenths were plated on M9 minimal media lacking histidine and supplemented with 20 μg/μl chloramphenicol. The plates were incubated at 37°C for one to three days before counting the colonies.

The two-hybrid assays used strain L40 (27) or its *sir* derivatives that express LexA-SIR4C and GAD-INC fusion proteins. A single colony was inoculated into 2 ml of SC-T-L liquid media and grown at 30°C for 24 hours. The yeast cells were spotted (ten-fold serial dilutions) onto solid SC-T-L-H media that was supplemented with 1-5 mM 3-amino 1,2,4-triazol (3-AT). As controls, cells were also spotted onto SC-T-L media. Plates were incubated at 30°C for three to four days before being imaged to record their growth.

RESULTS AND DISCUSSION

To test whether the interaction between Ty5 integrase (IN) and Sir4p is the primary determinant of Ty5 target site choice, integration was measured at DNA sites to which Sir4p is tethered (Figure 1A). The Sir4p C-terminus (SIR4C; amino acids 951-1358), which interacts withTy5 integrase in two-hybrid and *in vitro* binding assays (15), was expressed as a fusion protein with the LexA DNA binding domain (LexA-SIR4C). LexA-SIR4C was tethered to a target plasmid through LexA operators, which, in turn, were flanked by three to four kilobases of Arabidopsis DNA that serve as a landing site for Ty5 and prevent insertions from compromising plasmid function. To measure targeted integration, the plasmid was

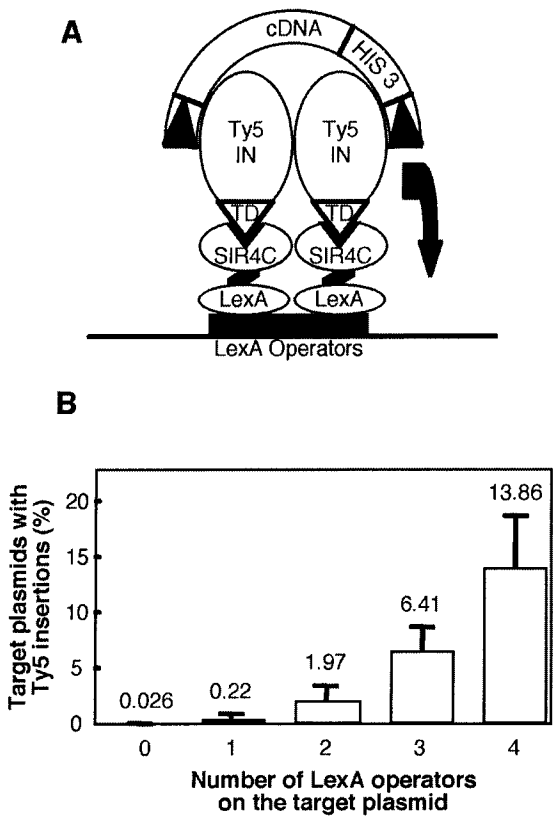

Figure 1. Ty5 integrates to sites of tethered Sir4p. (A) A tethered integration assay. The target plasmid carries LexA operators that bind LexA fusion proteins. LexA-SIR4C is shown interacting with the targeting domain (TD) of Ty5 integrase that is complexed to its cDNA. Arabidopsis DNA flanks the operators to create a Ty5 landing pad. The target plasmid also has selectable markers and replication origins that function in yeast (*TRP1*, 2m) and *E. coli* (chloramphenicol resistance gene, ColE1). Details of the tethered integration assay are described in the text. Briefly, after induction of Ty5 transposition, target plasmids are recovered in *E. coli* by selecting either for chloramphenicol resistance (Chl^r) or for chloramphenicol resistance and histidine prototrophy (Chl^rHis^+). The latter selection determines the number of target plasmids that carry Ty5 with its *HIS3* marker. The ratio of Chl^rHis^+ to Chl^r indicates the efficiency with which Ty5 integrates into the plasmid. (B) Monitoring integration to sites of tethered Sir4p. The efficiency of Ty5 integration to the target plasmid correlates with the number of LexA operators.

introduced into a yeast strain with a galactose-inducible Ty5 element (12). After growth on galactose, transposition events were selected by plating cells onto media lacking histidine. Ty5 carries a *his3AI* marker gene, and splicing of the Ty5 mRNA removes an inactivating intron, thereby reconstituting a functional *HIS3* gene upon reverse transcription and cDNA integration (28). Total DNA was prepared from His^+ yeast cells and used to transform a *hisB* *E. coli* strain. The *HIS3* gene within Ty5 complements the *E. coli hisB* mutation (29).

Because the target plasmid also carries a chloramphenicol resistance gene, plasmids with Ty5 insertions confer a Chl^rHis^+ phenotype to *E. coli*. The ratio of Chl^rHis^+ colonies (target plasmids with Ty5) to Chl^r colonies (target plasmids) measures the efficiency of integration to the target plasmid.

LexA-SIR4C created a strong Ty5 integration hotspot when tethered to the target plasmid. With four copies of the LexA operator, approximately 14% of the recovered target plasmids carried Ty5 insertions (Figure 1B). Targeting displayed a strict dependence on the number of LexA operators, suggesting that targeting efficiency is determined by the amount of SIR4C tethered to the plasmid. Ty5 insertions into the target plasmid were true integration events, as evidenced by target site duplications flanking several characterized insertions (data not shown).

Sir4p interacts with a number of proteins, including Sir2p and Sir3p, and loss of Sir proteins significantly decreases Ty5 target specificity (14). To test whether SIR4C requires other components of yeast heterochromatin for its interaction with IN, two-hybrid interactions were measured between IN and SIR4C in the absence of Sir proteins. These assays used LexA-SIR4C and a fusion protein generated between the Ty5 IN C-terminus (INC) and the Gal4p transcriptional activation domain (GAD) (15). The six amino acid targeting domain is located within INC and corresponds to positions 1092-1097 in the Ty5 polyprotein. The strength of the INC/SIR4C two-hybrid interaction was determined by expression of a *HIS3* reporter with upstream LexA operators (27). *HIS3* expression confers growth on media without histidine and with the inhibitor 3-amino 1,2,4-triazol (3-AT). The INC/SIR4C two-hybrid interaction was not significantly affected by loss of Sir2p, Sir3p or Sir4p (Figure 2A). In addition, SIR4C interacts with fusion proteins generated between GAD and nine amino acids of Ty5 integrase that encompass the six amino acid targeting domain (GAD-TD, Figure 2B). These results, coupled with previous *in vitro* binding studies (15), support the conclusion that INC and SIR4C interact directly through the six amino acid targeting domain.

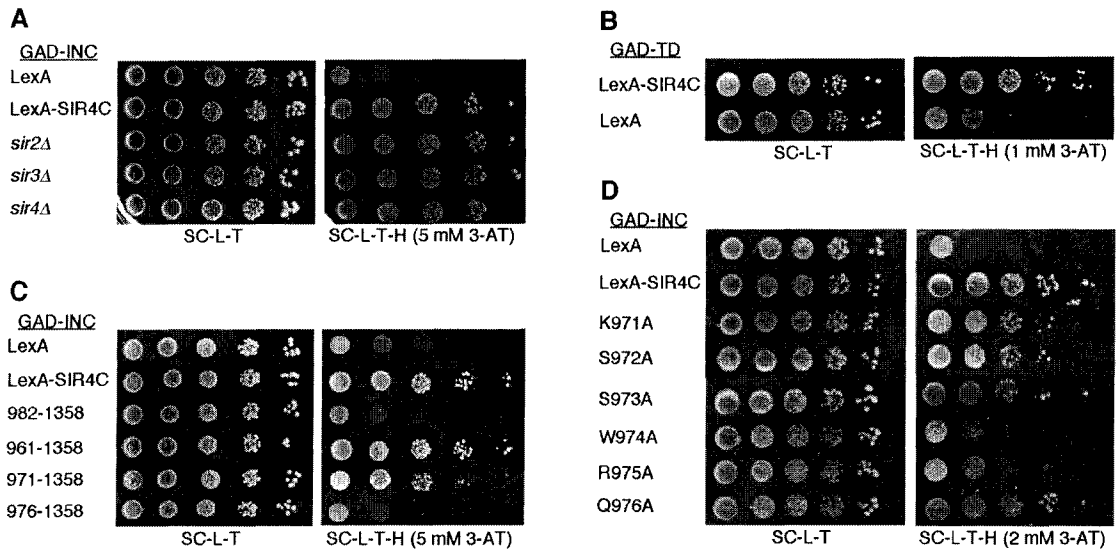

Figure 2. Defining the determinants that mediate integrase/Sir4p interactions. In panels A-D, two-hybrid assays measure the ability of LexA-SIR4C fusion proteins to interact with a fusion protein generated between the Gal4p activation domain (GAD) and Ty5 integrase. The GAD-INC construct has 258 amino acids from the C-terminus of Ty5 integrase. The GAD-TD construct has nine amino acids from integrase that encompass the six amino acid targeting domain. The strength of interaction is assayed by the expression of a *HIS3* reporter gene under the control of upstream LexA operators. Ten-fold serial dilutions of cells were spotted onto media that selects either for the plasmids expressing the fusion proteins (SC-L-T) or for both the plasmids and *HIS3* expression (SC-T-L-H plus the inhibitor 3-amino 1,2,4-triazol or 3-AT). (A) Sir-independence of the integrase/Sir4p interaction. Each of the *sir* strains expresses LexA-SIR4C. (B) Sufficiency of the targeting domain for Sir4p interactions. (C) The integrase-interacting domain of Sir4p. As indicated on the left-hand side of the panel, various N-terminal truncations of SIR4C were tested for their ability to interact with GAD-INC. Residues 971-975 of Sir4p are required for the INC/SIR4C interaction. (D) Amino acid residues of SIR4C required for the integrase/Sir4p interaction. Substitutions with alanine at tryptophan 974 (W974A) and arginine 975 (R974A) significantly disrupted the INC/SIR4C interaction.

To further map the region of SIR4C that interacts with INC, a series of SIR4C truncations were fused to the LexA DNA-binding domain (Figure 2C and data not shown). One SIR4C truncation (amino acids 982-1358) lost its ability to interact with INC (Figure 2C). This construct differs by only 30 amino acids from SIR4C constructs used in the previous experiments (amino acids 951-1358). A series of additional constructs were made with SIR4C N-termini corresponding to positions 961, 971, and 976. Of these, only the construct beginning at residue 976 failed to interact with INC, indicating that the region spanning residues 971 to 975 is critical for the INC/SIR4C interaction. To pinpoint essential residues, alanine was substituted at each of these amino acid positions. Only the W974A and

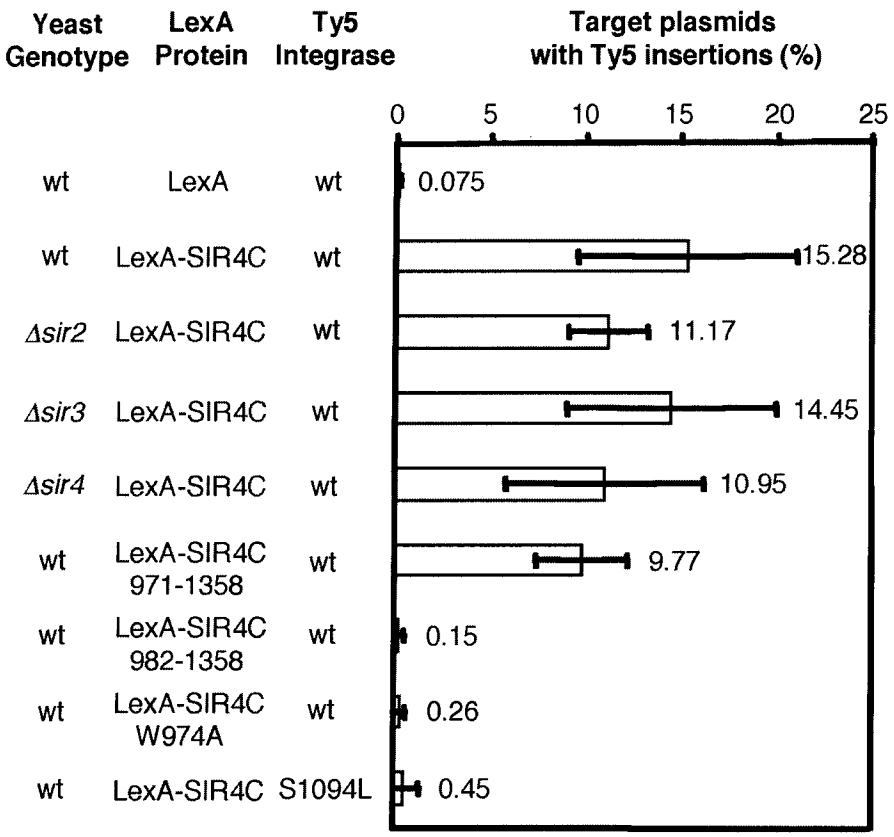

Figure 3. Tethered integration requires interaction between integrase and Sir4p. The tethered integration assay was performed with LexA-SIR4C in the absence of Sir2p, Sir3p and Sir4p. As with the two-hybrid assays, no effect on tethered integration was observed due to the absence of Sir proteins. Deletions or mutations in SIR4C that abrogate two-hybrid interactions (i.e. LexA-SIR4C 982-1358, W974A) prevent tethered integration. Similarly, mutations in the Ty5 the targeting domain (S1094L) that randomize chromosomal integration patterns (15, 16) do not support integration to sites of tethered SIR4C.

R975A substitutions disrupted the two-hybrid interaction (Figure 2D). In no case was failure of the two-hybrid interaction due to differences in expression of the various LexA-SIR4C constructs; all were expressed at comparable levels as measured by immunoblotting experiments performed with a LexA-specific antibody (data not shown).

We tested whether the requirements for the INC/SIR4C two-hybrid interaction correlated with requirements for targeted integration. This was invariably the case. Loss of

Sir proteins affected neither two-hybrid interactions nor targeting efficiency of Ty5 to sites of tethered SIR4C (Figure 3). When the various SIR4C derivatives were tested, only those fusion proteins that supported two-hybrid interactions created integration hotspots. For example, the construct beginning at residue 971 of Sir4p created an integration hotspot comparable to SIR4C, whereas no significant targeting was observed with the construct beginning at residue 982. Furthermore, the W974A *sir4* mutation, which abrogated two-hybrid interactions, also failed to target integration. In a complementary experiment, the tethered integration assay was performed with a Ty5 element with a mutation in its targeting domain (S1094L). This mutation randomizes genomic integration patterns and disrupts interactions with SIR4C in both two-hybrid and *in vitro* binding assays (15, 16). The Ty5 S1094L mutation also prevented integration to sites of tethered SIR4C (Figure 3).

Having defined the targeting determinants of integrase and its interacting partner, Sir4p, we next asked whether we could engineer Ty5 elements with altered target specificity. To accomplish this, we replaced the six amino acid Ty5 targeting domain with peptide motifs with known protein ligands. In one construct (Ty5-Rad9p), the targeting domain was swapped with a thirteen amino acid motif from Rad9p, which interacts with the two forkhead-associated domains of Rad53p (FHA1 and FHA2) (30, 31). In a second construct (Ty5-NpwBP), the targeting domain was replaced with a twelve amino acid, proline-rich motif from the human nuclear protein NpwBP (25). This motif interacts with the WW domain of a second nuclear protein, Npw38. The modified integrases were initially tested in two-hybrid assays with their respective protein partners, and interactions were comparable to the INC/SIR4C two-hybrid interaction (data not shown).

Neither of the targeting domain modifications compromised IN function, as transposition frequencies of Ty5-Rad9p and Ty5-NpwBP were comparable to those of a wild type element (data not shown). Remarkably, the efficiencies with which modified elements targeted to sites of tethered LexA-FHA1 and LexA-Npw38 were comparable to the efficiency with which the wild type element targeted to sites of tethered SIR4C (Figure 4A). Targeting required both the LexA operators and either the tethered Npw38 or the FHA1 domain. These results indicate that Ty5 target specificity can be altered, and suggest that

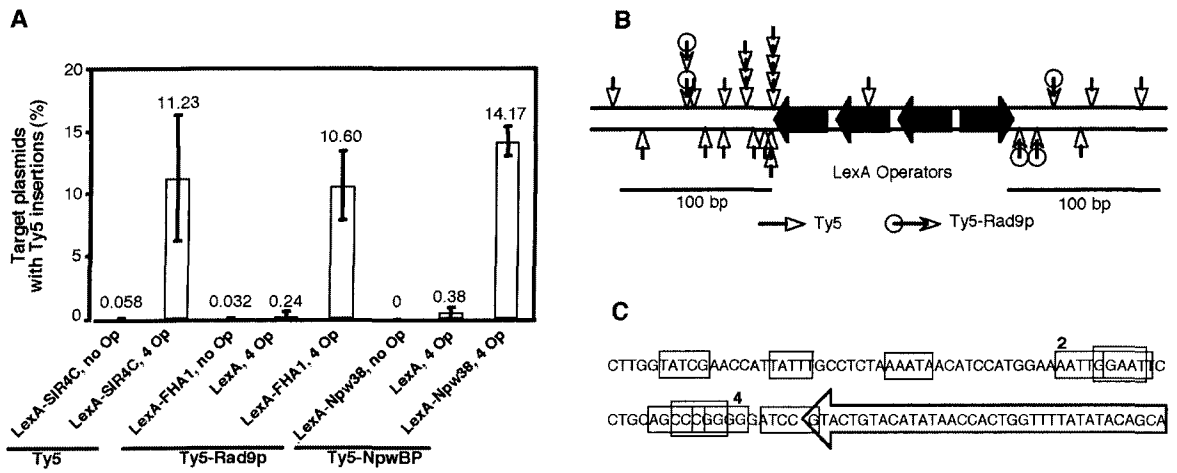

Figure 4. Altering specificity and precision of Ty5 integration. (A) Ty5-Rad9p and Ty5-NpwBP integrate at sites of tethered FHA1 and Npw38, respectively. For Ty5-Rad9p, the targeting domain of Ty5 was replaced with a peptide motif of Rad9p that interacts with the FHA1 domain of Rad53p (24). Similarly for Ty5-NpwBP, the targeting domain was replaced with a proline-rich motif of NpwBP that interacts with Npw38 (25). Targeting was measured to sites of tethered LexA-FHA1 and LexA-Npw38 and requires both the LexA operators and LexA fusion proteins. (B) Precision of integration to sites of tethered proteins. Twenty-six Ty5 insertions generated by both wild type and modified Ty5 elements were characterized by DNA sequencing. The insertions occurred within 120 bases of the LexA operators in the target plasmid. Ty5 insertions on the top strand are oriented 5' to 3'; insertions in the bottom strand are in the opposite orientation. (C) DNA sequence of 14 insertion sites adjacent to the leftmost LexA operator depicted in panel B. The sequence of the LexA operator is enclosed by an arrowhead. Target sites are boxed, and numbers adjacent to a given box denote the number of independent insertions that occurred at that site.

Ty5 variants with a range of integration specificity can be generated by substituting the targeting domain with peptide aptamers that recognize different chromosomal proteins.

To characterize where Ty5 integrates on the target plasmid, 26 insertions generated by both wild type Ty5 and Ty5-Rad9p were analyzed by DNA sequencing (Figure 4B, 4C). All 26 insertions occurred within 120 bases of the nearest LexA operator. No orientation preference was observed, and no obvious sequence consensus defined the insertion sites. The insertions that clustered near the leftmost LexA operator (as depicted in Figure 4B) displayed a regular periodicity of approximately 10 to 12 base pairs, suggesting that they occurred on the same face of the DNA helix. For 18 of the 26 insertions, DNA sequences were obtained from both ends of the element, and all 18 were flanked five base target site duplications.

The narrow integration window next to the LexA operators contrasts with the integration pattern observed for chromosomal Ty5 insertions, which typically occur within a three kilobase window flanking the *HM* silencers or the subtelomeric X repeats (12, 13). The tethered Ty5 integration patterns more closely resemble those of Ty3, which occur within 1-2 bases upstream of genes transcribed by RNA Pol III (5). Ty3 integration specificity requires transcription factor IIIB (9), and like LexA-SIR4C, TFIIIB occupies a well-defined chromosomal site. Collectively, these targeting patterns suggest that the precision of integration is determined primarily by the physical location of the protein or protein complex recognized by the integration machinery.

Retroviral vectors are widely used for gene delivery in gene therapy, in part because viral integration generates stable, defined, chromosomal insertions (32). The randomness of retroviral integration, however, is potentially hazardous and could have deleterious genetic effects, for example, by creating loss-of-function mutations or by activating oncogenes. A previous approach to control retroviral integration has been to fuse sequence-specific DNA binding domains to retroviral integrases (33-36). This approach has proven effective in *in vitro* integration assays, but because the integrase modifications often compromise viral replication, this approach has not been successfully utilized *in vivo*. The findings described here suggest an alternative approach for controlling retroviral integration, wherein retroviral integrases are modified to carry small peptide aptamers that recognize proteins bound to chromosomal target sites. In addition, the results have relevance for understanding eukaryotic genome organization. The successful proliferation of retrotransposons is thought to be due to their ability to identify safe-havens in the genome where integration is not harmful to their hosts (3, 4). The widespread association of eukaryotic retrotransposons with heterochromatin suggests that these gene-poor domains are one such safe-haven (3). If Ty5's strategy for selecting integration sites is employed by other retrotransposons, then targeted integration may have significantly shaped eukaryotic genome organization.

ACKNOWLEDGEMENTS

This study is dedicated to the memory of Francis J. Voytas (father of D.F.V.). The work was supported by grants to D.F.V. from the American Cancer Society (RPG9510106MBC) and the National Institutes of Health (GM61657).

REFERENCES

1. Medstrand, P., Van De Lagemaat, L. N. & Mager, D. L. (2002) *Genome Res.* **12,** 1483-1495.

2. SanMiguel, P., Tikhonov, A., Jin, Y. K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P. S., Edwards, K. J., Lee, M., Avramova, Z. & Bennetzen, J. L. (1996) *Science* **274,** 765-768.

3. Boeke, J. D. & Devine, S. E. (1998) *Cell* **93,** 1087-1089.

4. Craig, N. L. (1997) *Annu. Rev. Biochem.* **66,** 437-474.

5. Chalker, D. L. & Sandmeyer, S. B. (1992) *Genes Dev.* **6,** 117-128.

6. Devine, S. E. & Boeke, J. D. (1996) *Genes Dev.* **10,** 620-633.

7. Kirchner, J., Connolly, C. M. & Sandmeyer, S. B. (1995) *Science* **267,** 1488-1491.

8. Aye, M., Dildine, S. L., Claypool, J. A., Jourdain, S. & Sandmeyer, S. B. (2001) *Mol. Cell. Biol.* **21,** 7839-7851.

9. Yieh, L., Kassavetis, G., Geiduschek, E. P. & Sandmeyer, S. B. (2000) *J. Biol. Chem.* **275,** 29800-29807.

10. Yieh, L., Hatzis, H., Kassavetis, G. & Sandmeyer, S. B. (2002) *J. Biol. Chem.* **277,** 25920-25928.

11. Zou, S., Wright, D. A. & Voytas, D. F. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 920-924.

12. Zou, S., Ke, N., Kim, J. M. & Voytas, D. F. (1996) *Genes Dev.* **10,** 634-645.

13. Zou, S., Kim, J. M. & Voytas, D. F. (1996) *Nucleic Acids Res.* **24,** 4825-4831.

14. Zhu, Y., Zou, S., Wright, D. & Voytas, D. (1999) *Genes Dev.* **13,** 2738-2749.

15. Xie, W., Gai, X., Zhu, Y., Zappulla, D. C., Sternglanz, R. & Voytas, D. F. (2001) *Mol. Cell. Biol.* **21,** 6606-6614.

16. Gai, X. & Voytas, D. F. (1998) *Mol. Cell* **1,** 1051-1055.

17. Schroder, A., Shinn, P., Chen, H., Berry, C., Ecker, J. & Bushman, F. (2002) *Cell* **110,** 521-529.

18. Kalpana, G. V., Marmon, S., Wang, W., Crabtree, G. R. & Goff, S. P. (1994) *Science* **266,** 2002-2006.

19. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1987) *Current Protocols in Molecular Biology* (Greene/Wiley Interscience, New York).

20. Estojak, J., Brent, R. & Golemis, E. A. (1995) *Mol. Cell. Biol.* **15,** 5820-5829.

21. Ke, N. & Voytas, D. F. (1997) *Genetics* **147,** 545-556.

22. James, P., Halladay, J. & Craig, E. A. (1996) *Genetics* **144,** 1425-1436.

23. Gao, X., Rowley, D. J., Gai, X. & Voytas, D. F. (2002) *J. Virol.* **76,** 3240-3247.

24. Liao, H., Yuan, C., Su, M. I., Yongkiettrakul, S., Qin, D., Li, H., Byeon, I. J., Pei, D. & Tsai, M. D. (2000) *J. Mol. Biol.* **304,** 941-951.

25. Komuro, A., Saeki, M. & Kato, S. (1999) *J. Biol. Chem.* **274,** 36513-36519.

26. Christianson, T. W., Sikorski, R. S., Dante, M., Shero, J. H. & Hieter, P. (1992) *Gene* **110,** 119-122.

27. Hollenberg, S. M., Sternglanz, R., Cheng, P. F. & Weintraub, H. (1995) *Mol. Cell. Biol.* **15,** 3813-3822.

28. Curcio, M. J. & Garfinkel, D. J. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 936-940.

29. Struhl, K. & Davis, R. W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5255-5259.

30. Sun, Z., Hsiao, J., Fay, D. S. & Stern, D. F. (1998) *Science* **281,** 272-274.

31. Durocher, D., Henckel, J., Fersht, A. R. & Jackson, S. P. (1999) *Mol. Cell* **4,** 387-394.

32. Verma, I. M. & Somia, N. (1997) *Nature* **389,** 239-242.

33. Goulaouic, H. & Chow, S. A. (1996) *J. Virol.* **70,** 37-46.

34. Katz, R. A., Merkel, G. & Skalka, A. M. (1996) *Virology* **217,** 178-190.

35. Bushman, F. D. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 9233-9237.

36. Bushman, F. D. & Miller, M. D. (1997) *J. Virol.* **71,** 458-464.

CHAPTER 3. PHOSPHORYLATION REGULATES TY5 INTEGRATION SPECIFICITY IN *SACCHAROMYCES CEREVISIAE*

A manuscript to be submitted to *Molecular Cell*

Junbiao Dai[1], Weiwu Xie[2], Troy Brady[3] and Daniel F. Voytas[4]

ABSTRACT

A critical step in the retroelement life cycle is the integration of cDNA into the host chromosome. The selection of integration sites is not random, and to avoid the deleterious consequences of integration, 1) the host may evolve ways to regulate integration specificity, 2) the element may actively select target sites, or 3) both mechanisms may be used. The yeast Ty5 retrotransposon integrates specifically into silent chromatin (heterochromatin) at the telomeres and silent mating loci. Targeting is mediated by the interaction between Ty5 integrase and the silent information regulator 4 (Sir4p). A small domain at the C-terminus of integrase (the targeting domain, LDSSPP) is critical for target specificity. Here we provide genetic and biochemical evidence that the targeting domain is post-translationally modified by phosphorylation. MS/MS data indicate that the second serine (S1095) within the targeting domain is phosphorylated. As indicated by surface plasmon resonance spectroscopy, the phosphorylation of S1095 is required for interaction with Sir4p. A collection of non-lethal kinase deletion mutants was screened for kinases that effect Ty5 target specificity. Both a wild type Ty5 element and one with a S1095E mutation (which mimics phosphorylation) were used in the screen. Among the kinases found to effect targeting was the checkpoint kinase, Dun1p. In Ddun1 strains, transposition was completely abolished; however, Dun1p appears to affect targeting indirectly, because purified Dun1p failed to phosphorylate the targeting domain *in vitro*. Collectively, the data support a model in which both the element – by encoding targeting determinants – and the host – by post-translational modification –

---

[1] Primary researcher and author

[2] Second author who carried out the screening of the kinase deletion strains, *in vitro* binding of GBD/GST-TD to Sir4p and phosphatase treatment experiment

[3] Third author who carried out the *in vitro* tests of the interaction between TD and Sir4C

[4] Professor and corresponding author, Department of Genetics, Development and Cell Biology, Iowa State University, Ames, IA 50011

determine target site choice. Furthermore, multiple kinases, such as Dun1p, play an indirect role in target specificity, perhaps by modulating the chromatin targets of Ty5.

INTRODUCTION

The ability of retroviruses to insert their cDNA into the host genome has been harnessed to deliver therapeutic genes for human gene therapy (Bushman et al., 2005) (Kootstra and Verma, 2003; Marshall, 2001; Sandoval Rodriguez et al., 2005). Gene transfer using retroviruses, however, can be potentially problematic, since insertions can alter the expression or function of host genes and lead to diseases such as cancer (Check, 2002; Kimmelman, 2005; Kootstra and Verma, 2003; Marshall, 2001). Studies have shown that retroviral insertions display nonrandom distribution patterns on chromosomes (Mitchell et al., 2004; Schroder et al., 2002; Wu et al., 2003). Transcription units, especially active genes, are favored integration sites for human immunodeficiency virus (HIV) (Mitchell et al., 2004; Schroder et al., 2002; Wu et al., 2003). Similar patterns were revealed for SIV integration into macaque chromosomes (Hematti et al., 2004), suggesting that targeted integration into active genes is a common feature of lentiviruses and possibly occurs by a similar targeting mechanism. Differing from HIV and SIV, murine leukemia virus (MLV) shows preferred integration near transcription start sites (Mitchell et al., 2004). A model has emerged (the tethered integration model) in which target specificity is explained by interaction between pre-integration complexes (PIC) and DNA bound proteins such as transcription factors (Bushman, 2003). Understanding the mechanism of targeted integration clearly has important implications for improving gene therapy.

Target specificity is particularly pronounced for the five retrotransposon families of budding yeast (Ty1 - Ty5) (Lesage and Todeschini, 2005). Ty3 integrates specifically into the transcription initiation sites of genes transcribed by RNA pol III, including tRNA, 5S and U6 genes. Integration sites are restricted within a few nucleotides of the start of transcription (Chalker and Sandmeyer, 1992). The underlying mechanism has not been determined, but it appears that it does not occur through recognition of primary DNA sequences. Rather, and consistent with the tethered integration model, targeting is likely mediated by the interaction between Ty3 PIC and proteins bound to transcription initiation sites. In support of this, it has

been shown that pol III transcription factors (TFIIIB and TFIIIC) are required for the position-specific integration of Ty3 (Kirchner et al., 1995).

Ty1 shares a similar integration pattern with Ty3 and also targets to pol III transcribed genes. However, Ty1 targeting is not as precise as Ty3. Rather, Ty1 integrates into a window spanning about 600bp upstream of the transcription initiation sites(Boeke and Devine, 1998; Devine and Boeke, 1996). Although both Ty1 and Ty3 insert into similar regions, their targeting mechanism may not be the same, as suggested by their significantly different integration patterns. Little is known about Ty2 and Ty4 targeting except that they are also associated with pol III transcribed genes and share the same pattern as Ty1 (Kim et al., 1998).

Ty5 displays a very different integration specificity from the other Tys. About 95% of the endogenous and *de novo* Ty5 insertions occur in or are associated with regions of yeast heterochromatin. These regions include the telomeres and the silent mating loci (*HML* and *HMR*) (Zou et al., 1996; Zou and Voytas, 1997). The targeting determinant of Ty5 has been mapped to the C-terminus of Ty5 integrase. A single amino acid substitution within the targeting domain (TD, LDSSPP) is able to randomize Ty5 integration (Gai and Voytas, 1998; Xie et al., 2001). In addition, it has been shown that targeted integration of Ty5 to heterochromatin occurs via interaction between the integrase targeting domain and the silent information regulator 4 (Sir4p) (Xie et al., 2001; Zhu et al., 2003). Furthermore, targeting specificity of Ty5 can be altered by swapping TD with exogenous motifs (Zhu et al., 2003). The tethered targeting model is clearly valid for Ty5, because integration specificity is determined solely by protein-protein interactions (i.e. the integrase/Sir4p interaction).

In this study, we investigate the role of post-translational modification of integrase in controlling Ty5 target specificity. We report that the targeting domain of Ty5 integrase is phosphorylated and that phosphorylation is necessary for productive interactions with Sir4p. In addition, we show that the DNA damage checkpoint kinase, Dun1p, participates in regulating Ty5 transposition and targeting either directly or indirectly. Our results suggest

that the host cell – through post-translational modification – plays a critical role in determining how transposable elements affect the integrity of its genome.

RESULTS:

In vitro binding of TD to Sir4p is blocked by treatment with phosphatase

The Ty5-encoded targeting domain (TD) interacts with the Sir4p C-terminus (Sir4C) and tethers the pre-integration complex to heterochromatic regions giving rise to Ty5's insertion preference (Xie et al., 2001; Zhu et al., 2003). TD is a six amino acid motif: **L**DS**SP**P. The four residues in bold, including the two serines, denote amino acids that are required for targeting to heterochromatin. We have shown that when TD is expressed as a fusion protein with the Gal4p DNA binding domain (GBD) it interacts with Sir4C that is expressed *in vitro* (Figure 1A, (Xie et al., 2001)). In these experiments, the GBD-TD fusion protein was expressed in yeast cells and immunopurified using a GBD antibody; however, when TD was fused to GST and expressed and purified from *E. coli*, it could not bind to Sir4C (Figure 1A left panel). It is possible that the different protein scaffolds (i.e. GBD vs. GST) affect TD's function, although this was not the case when TD was expressed in yeast as a fusion to the Gal4p activation domain (GAD) or LexA (Xie et al., 2001). Alternatively, the difference in binding may due to the fact that TD was purified from yeast versus *E. coli*. We hypothesized that TD expressed in yeast cells is post-translationally modified, perhaps by phosphorylation of one or more of the serines. To test this hypothesis, the immunopurified TD fusion protein was treated with the λ protein phosphatase before mixing with the Sir4C protein. Binding of the treated protein was reduced to the background level (Figure 1A).

Amino acid substitutions suggest serine phosphorylation of TD.

Amino acid substitutions in critical residues of TD, even conservative changes such as L1092V, cause serious Ty5 targeting defects (about a 20-fold decrease in targeting using a plasmid-based targeting assay) (Gai and Voytas, 1998; Xie et al., 2001). Despite this sensitivity to mutation, we reasoned that the serines in TD might be able to be substituted by threonine without affecting TD function, particularly if TD is indeed phosphorylated *in vivo*

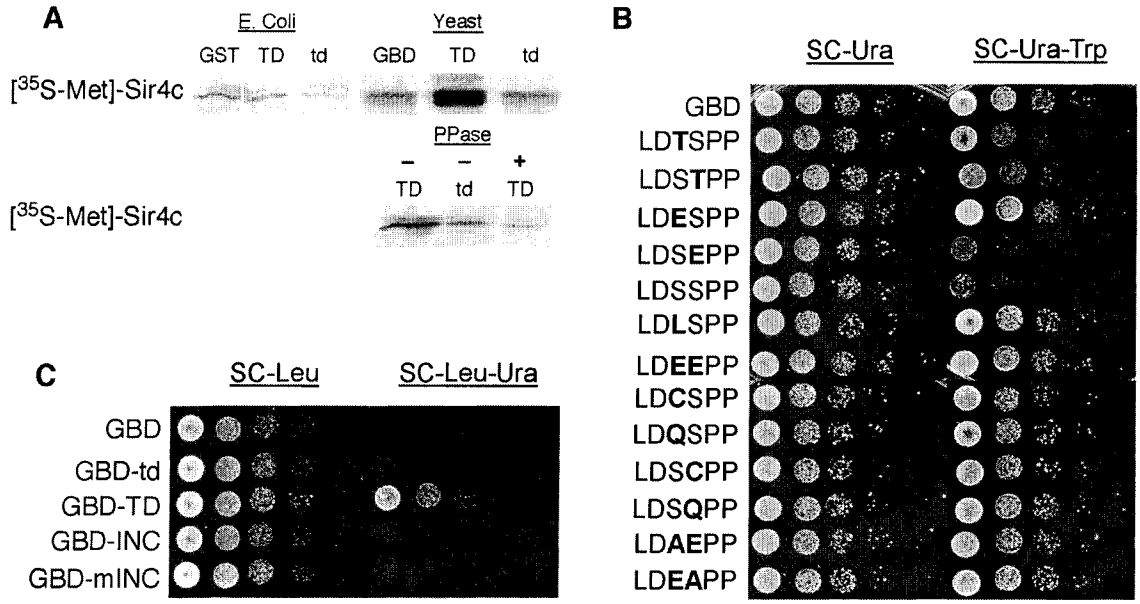

Figure 1. Post-translational modification of the Ty5 targeting domain (TD). (A) *In vitro* interaction between Sir4C and TD. Sir4C interacts with TD purified from yeast but not *E. coli*. Sir4C was expressed and labeled with [^{35}S]methionine by coupled transcription and translation. GBD-TD and GBD-td were expressed in yeast and immunoaffinity purified with anti-GBD agarose beads. GST-TD and GST-td were expressed in *E. coli* and purified with glutathione agarose beads. The interaction between Sir4C and TD is abolished when TD is treated with phosphatase. The immunopurified TD fusion protein was treated with the l protein phosphatase at 30°C for 30 min before mixing with the Sir4C protein. (B) Silencing was established when GBD-TD were tethered to the weakened *HMR* locus at a triple UAS_G site. Serial, 10-fold dilutions of cells were plated onto control (SC-Ura) or test (SC-Ura-Trp) media to measure silencing of the *TRP1* reporter gene at *HMR*. Different td mutations were also tested. The mutated residues are bold. (C) Overexpression of GBD-TD, but not GBD-td, GBD-INC or GBD-mINC, disrupts telomeric silencing. In the test strains (UCC3505), the *URA3* gene is located at telomeres VIIL. *URA3* expression was measured by growth of the yeast cells on media lacking uracil. TD, LDSSPPNTS; td, LDLSPPNTS

by certain serine/threonine kinases. In addition, phosphorylation may be mimicked by substituting the serines with a negatively charged amino acid (e.g. glutamic acid). Mutant versions of TD were generated and tested for interaction with Sir4C and their effects on target specificity.

The interaction between TD and Sir4C can be measured *in vivo* by the ability of GBD-TD fusion proteins to reconstitute transcriptional silencing at a crippled *HMR* locus (tethered silencing) (Xie et al., 2001). In this assay, the *HMR*-E transcriptional silencer is rendered non-functional by mutation, and as such, a *TRP1* marker gene inserted at *HMR* is

expressed. (Chien et al., 1993). Inserted into the crippled *HMR*-E silencer is a sequence that binds GBD (UAS_G), thereby allowing various GBD fusion proteins to be tested for their affect on transcription of the adjacent *TRP1* gene. We previously found that expression of a GBD-TD fusion protein silences *TRP1* in the tethered silencing assay, presumably because it interacts with Sir4p and establishes silent chromatin (Xie et al., 2001). Here we respectively substituted each serine of TD in the GBD fusions by either threonine or glutamic acid. The various mutants were expressed in yeast and silencing of *TRP1* at *HMR* was measured by plating serial dilutions of yeast cells onto non-selective and selective media. Wild type TD silenced TRP1 at least 100 times more effectively than the control (GBD only) and the S1094L mutant (Figure 1B). The S1095E mutant was as effective in silencing as the wild type, and the threonine substitutions had an intermediate phenotype, indicating that they cannot totally substitute for the native serines. Furthermore, substitutions with cysteine, which shares a similar structure with serine, failed to nucleate silencing. Therefore, a functional hydroxyl group is required for TD function. Taken together, these results suggest that TD is modified by kinase(s) and the modification is required for function. The data also implicate the second serine as the site of phosphorylation, since neither S1094E nor the S1094E, S1095E double mutant nucleated silencing.

Our lab has previously shown that when GBD-TD is overexpressed, telomeric marker genes that are normally silenced become expressed (Xie et al., 2001). This loss of silencing is due to the degradation of Sir4p caused by the Sir4p/TD interaction (Fuerst and Voytas, manuscript in preparation). We tested the new mutants for their ability to disrupt telomeric silencing; however, unlike wild type TD, no anti-silencing activity was observed (data not shown). Loss of telomeric silencing appears to be unique to GBD-TD expression, because expression of the IN C-terminus, which includes TD, also fails to break silencing, even though other assays indicate that the IN C-terminus/Sir4p interaction is very robust. Therefore, loss of transcriptional silencing by this assay may not be a reliable indicator of TD/Sir4p interactions.

Interactions between TD and Sir4C can also be measured by two-hybrid assays (Xie et al., 2001). We tested S1094E, S1095E and the double mutant in two-hybrid assays, and

Table 1. Targeting assay results of Ty5 mutants.

Ty5 mutants	Sequence of TD	Targeting to HMR (%)
pNK254(WT)	LDSSPP	6.07±0.79
pWW79	LDSEPP	5.51±0.84
pWW102	LDESPP	0.70±0.72
pWW103	LDEEPP	1.37±0.90
pXW137	LDLSPP	1.13±0.66
pJB171	LDCSPP	2.72±2.03
pJB170	LDSCPP	2.70±1.20
pJB167	LDSAPP	1.56±0.86
pJB 169	LDAAPP	3.65±2.64

none of the mutants were found to interact with Sir4C (data not shown). This contrasts with the tethered silencing assay, in which the S1095E mutant was able to establish silencing at levels equivalent to wild type. It is possible that the negatively charged amino acid substitutions do not exactly mimic the phosphorylation. Alternatively, the interaction with Sir4C in the two-hybrid assay may differ from the interaction with the full-length Sir4p, which is measured in the tethered silencing assay.

To examine the effect of the E substitutions on Ty5 target specificity, several Ty5 mutants, including S1094E, S1095E and the double mutant, were constructed and tested using our plasmid-based targeting assay (Gai and Voytas, 1998). This assay monitors integration of Ty5 to a plasmid-borne *HMR* locus. For wild type Ty5, approximately 8 to 10% of transposition events occurred on the plasmid, whereas targeting defective mutants transpose to the plasmid at frequencies ranging from 0.5 to 3% (Xie et al., 2001). Consistent with previous data, we found that wild type Ty5 targets to the plasmid at a frequency approximating 6% (Table 1). The S1094E mutation had almost no effect on Ty5 target specificity, whereas the S1095E mutation decreased Ty5 targeting to 0.7±0.72%. The double mutation also had a severe effect on targeting (1.37±0.90%). Similarly, the S1095C and S1095A substitutions decreased targeting efficiency from about 6% to 2.70±1.20% and 1.56±0.86%, respectively. These results are consistent with the phenotypes observed for the

Table 2. MS/MS analysis of TD/td phosphorylation.

Constructs	P- or not?	Sites
$LDSSPPNTSH_6$	Yes	S1095
$LDLSPPNTSH_6$	Yes	S1095
$H_6LDSSPPNTS$	Yes	S1095
$LDSAPPNTSH_6$	No*	—

*When mutated to alanine, there is a trace amount of phosphorylation on S1094, but it is missing in all other constructs.

tethered silencing assay. This data supports the hypothesis that S1095 is phosphorylated and that phosphorylation can be mimicked by glutamic acid substitution.

TD is phosphorylated in vivo.

To test directly whether TD is phosphorylated and to determine the precise modification sites, TD was purified from yeast cells and subjected to mass spectrometry. For these experiments, a six-histidine tag (His6) was added to the N- and C-termini of TD to facilitate purification by nickel chelate chromatography. The His6-tagged TDs were first tested to ensure that the tags did not compromise function. This was accomplished by fusing the tagged domains to GAD and testing for interactions with Sir4C by the two-hybrid assay. As shown in Figure 2A, the His6-tagged TDs showed interactions of similar strength to the untagged TD (compare row 2 to row 4). The His6-tagged TD was then fused to glutathione S-transferase (GST), expressed in yeast and purified with Ni-NTA magnetic beads. The eluted protein was subjected to a second round of affinity chromatography using glutathione agarose. The nine-amino acid peptide, along with either the N- or C-terminal His6-tag, was cleaved from GST by factor Xa and concentrated with Ni-NTA magnetic beads. The purified 15-amino acid peptides were analyzed on a Q-TOF MS/MS system.

The MS spectrum indicated that there is a single phosphorylation site within TD (Figure 2B). The MS/MS data are summarized in Table 2, and consistent with the genetic data, S1095 was phosphorylated. Although genetic data indicate that S1094 is

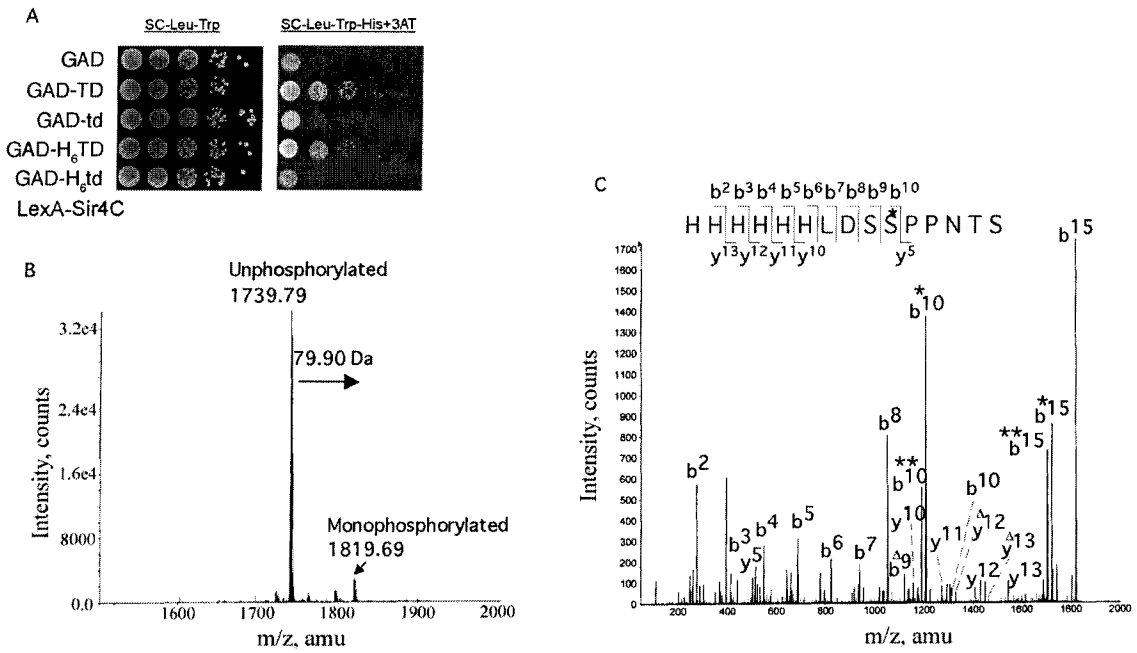

Figure 2. TD is phosphorylated *in vivo*. (A) Two-hybrid assays were conducted to test interactions between Sir4C and His_6-tagged TD. The addition of His_6 at the N-terminus of TD does not interfere TD to interact with Sir4C. (B) Deconvoluted mass spectrum of His_6-tagged TD. Values above each peak are the observed mass of the most abundant isotopic species. The observed mass difference between unphosphorylated TD and monophosphorylated TD corresponds closely to the calculated monoisotopic mass of HPO_3, which is 79.996 Da. (C) The ion fragmentations from tandem mass spectrum of the ion fragment corresponding to monophosphorylated TD (m/z 1819.69 in (B)). More than 70% of the observed peaks could be assigned when serine marked with * was considered to be the phosphorylation site. b* and b** are produced by the loss of HPO_3 and one or two H_2O from b ions. b^{Δ} and y^{Δ} ions are produced by the loss of one H_2O from b ions and y ions, respectively.

required for TD function, neither S1094 nor the downstream serine (S1099) or threonine (T1100) was phosphorylated. We also purified an S1095A TD variant and found that S1094 was phosphorylated in the mutant; however, the phosphorylation was only barely detectable even though the same expression and purification conditions were used. We therefore believe that S1094 is not phosphorylated *in vivo* or if it is, only at very low levels or perhaps only when S1095 (the preferred site) is mutated. Interestingly, the S1094L mutant is unambiguously phosphorylated at S1095. The genetic analysis indicated that S1094L greatly reduces Ty5 target specificity and fails to interact with Sir4C, yet it seems that this defect is not due to a change in TD phosphorylation.

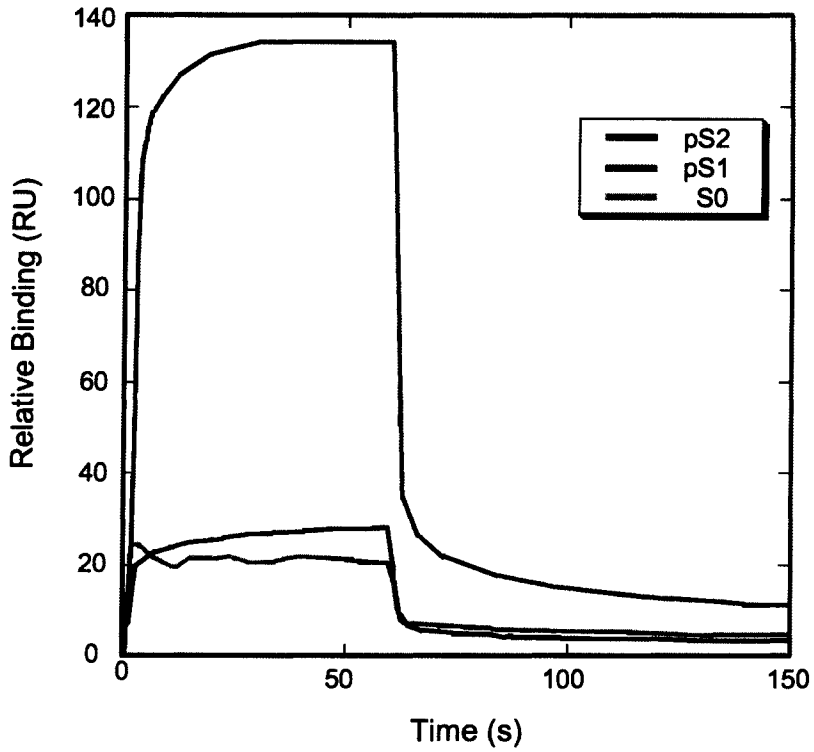

Figure 3. Surface plasmon resonance demonstrates the interaction between TD and Sir4C. S0, synthetic TD without phosphorylation; pS1, synthetic TD with phosphorylated S1094; pS2, synthetic TD with phosphorylated S1095. A significant association can only be detected when the second serine (S1095) within TD is phosphorylated. 2.5mM of peptides are used in the experiment.

Instead, it might be possible that the mutation causes a conformational change that blocks interaction with Sir4C. Alternatively, it might reduce S1095 phosphorylation. In support of this, we used a phospho-TD specific antibody (unpublished), and found that the S1094L mutant lowered phosphorylation at S1095 (Dai and Voytas, unpublished data). Further experiments will be required to clarify the significance of this observation.

Phosphorylation is required for TD to interact with Sir4C.

The genetic data suggest that phosphorylation is required for TD function. Furthermore, the S1095C substitution – a very subtle mutation – abolishes TD activity (Figure 1B). However, it is nonetheless possible that the mutations cause a structural change in the protein that secondarily leads to loss of TD function. Furthermore, in S1094 mutants,

S1095 is phosphorylated, indicating that if phosphorylation is necessary for binding, it is not sufficient.

To directly test whether S1095 phosphorylation is required for the interaction with Sir4C, an *in vitro* assay was performed using surface plasmon resonance (SPR) spectroscopy to monitor associations between unphosphorylated or phosphorylated TD (pTD) and GST-Sir4C. Three peptides were synthesized that include three nascent residues immediately up- and downstream of TD. One peptide was unphosphorylated (S0,) and two had a phosphoserine at either S1094 (pS1) or S1095 (pS2). This allowed us to test the importance of both phosphorylation and the position of the phosphate group for Sir4C interaction. In SPR experiments, S0 and pS1 exhibited very weak binding affinity, with the phosphate group on the first serine enhancing interaction approximately 2 to 3-fold. The weakness of the pS1 interaction is evidenced by the fact that its binding to GST-Sir4C never reached saturation. Phosphorylation of the second serine, however, increased TD affinity for Sir4C over 12-fold. These data not only show that phosphorylation is important for TD interaction with Sir4C *in vitro,* but also that the phosphate-enhanced interaction is specific for modifications on the second serine. While these data demonstrate a direct interaction between these two protein fragments, the binding of pS2 to the GST-Sir4C fusion under these conditions required high concentrations of peptide, indicating the affinity is weak. We speculate that high-affinity binding would require other portions of the IN C-terminus that extend beyond the 12 amino acid residues tested.

Multiple kinases affect Ty5 target specificity.

Since TD is phosphorylated, we sought to identify the responsible kinase(s). There are 135 kinases encoded by the yeast genome (Hunter et al., 2000), out of which 109 kinase knockout strains are available from the Saccharomyces Genome Deletion Project (Winzeler et al., 1999). Our plasmid-based targeting assay requires that the host strain is deficient in homologous recombination (Gai and Voytas, 1998), and so we needed to delete RAD52 in each of the kinase knockout strains. This was accomplished by first mating a *rad52*Δ strain to each knockout strain. Diploids were sporulated, and the resulting haploids selected that had both the kinase knockout and the *rad52*Δ allele. Five kinase knockout strains were

unable to mate with the *rad52*Δ strain (121E10, *ste20*Δ; 125F7, *fus3*Δ; 131B9, *pkh2*Δ; 139A8, *ste7*Δ; 149G7, *ste11*Δ; 143E10, *ctk1*Δ), two strains failed to sporulate (123G5, *pho85*Δ; 147G3, *sps1*Δ), and four strains failed to obtain desired haploid offspring (133H6, *bck1*Δ; 143C1, *dbf2*Δ; 171C10, *vps15*Δ; 171D4, *bud32*Δ).

Targeted transposition was tested in the remaining 97 kinase knockout strains as well as wild type controls of both the a and α mating type. Twenty-eight strains had targeting frequencies of less than 5%, and these frequencies were confirmed in at least two, and more typically three, independent tests. Among the 28 strains, three strains had targeting frequencies less than 2% (101E6, *kin3*Δ; 103A11, *yml059c*Δ; 117B8, *hsl1*Δ) and three strains had frequencies between 2% and 3% (114F11, *yck1*Δ; 129D8, *prk1*Δ; 145B9, *fab1*Δ). These frequencies approximate those observed for the Ty5 S1094L mutant.

It is unlikely that all the kinases we identified with targeting defects phosphorylate Ty5 TD. To further screen the mutants, we tested whether TD -- when purified from the six strains with the lowest targeting frequencies -- was able to interact with Sir4C. As described above, GBD-TD was expressed and purified from yeast and tested for its ability to pulldown radiolabeled Sir4C. We anticipated that based on the targeting assay data, binding would be abrogated, however no obvious differences were observed for any of the kinase knockout strains (data not shown).

The 28 kinase knockout strains with targeting efficiencies below 5% were subjected to an additional round of screening. As described above, S1095 is phosphorylated and the S1095E mutation can mimic phosphorylation (Figure 1B). We reasoned, therefore, that the S1095E substitution would bypass the requirement for TD phosphorylation and restore Ty5 targeting in kinase mutant backgrounds. A Ty5 plasmid bearing the S1095E substitution was transformed into the 28 kinase knockout strains and targeting frequencies determined. Targeted transposition of the S1095E mutant in 20 of the strains still occurred at low frequencies (data not shown). This could be because the kinases affect other features required for targeting, such as heterochromatin. In support of this, we found several kinases that affect telomeric transcriptional silencing. Characterization of these kinases will be

reported elsewhere. For some kinases, however, it appears the plasmid-based targeting assay is not a good measure of target specificity. This is supported by some additional experiments performed with strain 101E6 (*kin3*Δ). Eight chromosomal Ty5 insertions were recovered from the *kin3*Δ *strain,* and all were located in preferred Ty5 targets: five in telomeric regions and three at the *HM* loci (data not shown). Therefore for *kin3*Δ and perhaps some other kinase knockout strains, the plasmid-based targeting assay was not a reliable measure of target specificity.

Eight strains (110G12; *kin1*Δ; 115F2, *sat4*Δ; 117F6, *tpk3*Δ; 119D2, *mkk1*Δ; 124A9, *Irs1*Δ; 136F12, *prr2*Δ; 137D4, *chk1*Δ; 138F2, *dun1*Δ) were restored by the S1095E mutation in Ty5. In these strains, targeting frequencies of the S1095E mutant ranged from 6% to 12% (Table 3). Two strains (124A9, *Irs1*Δ; 138F2, *dun1*Δ) had consistent results in at least three individual tests, but the others showed considerable variation, with percentages ranging from less than 7% to greater than 10%. All, however, are possible candidates for either directly or indirectly modifying S1095.

Dun1p affects Ty5 transposition and the interaction between Sir4C and TD

As described above, the interaction between TD and Sir4p was observed *in vivo* using the tethered silencing, loss of telomeric silencing, and two-hybrid assays. Based on our *in vitro* experiments, which demonstrated a requirement for phosphorylation for the Sir4C/TD interaction, we expected that these assays would be affected in kinase knockouts that alter TD phosphorylation. To test this, we knocked out the eight kinases in the strains suitable for measuring *in vivo* interactions between Sir4C and TD.

Among the eight kinases tested, Dun1p, the DNA damage checkpoint kinase (Hunter et al., 2000), showed a strong effect on GBD-TD-induced loss of telomeric silencing and in the two-hybrid (Figure 4). However, no obvious effect was observed for the loss of Dun1p in the tethered silencing assay. This conflicting latter result may be explained if the effect of

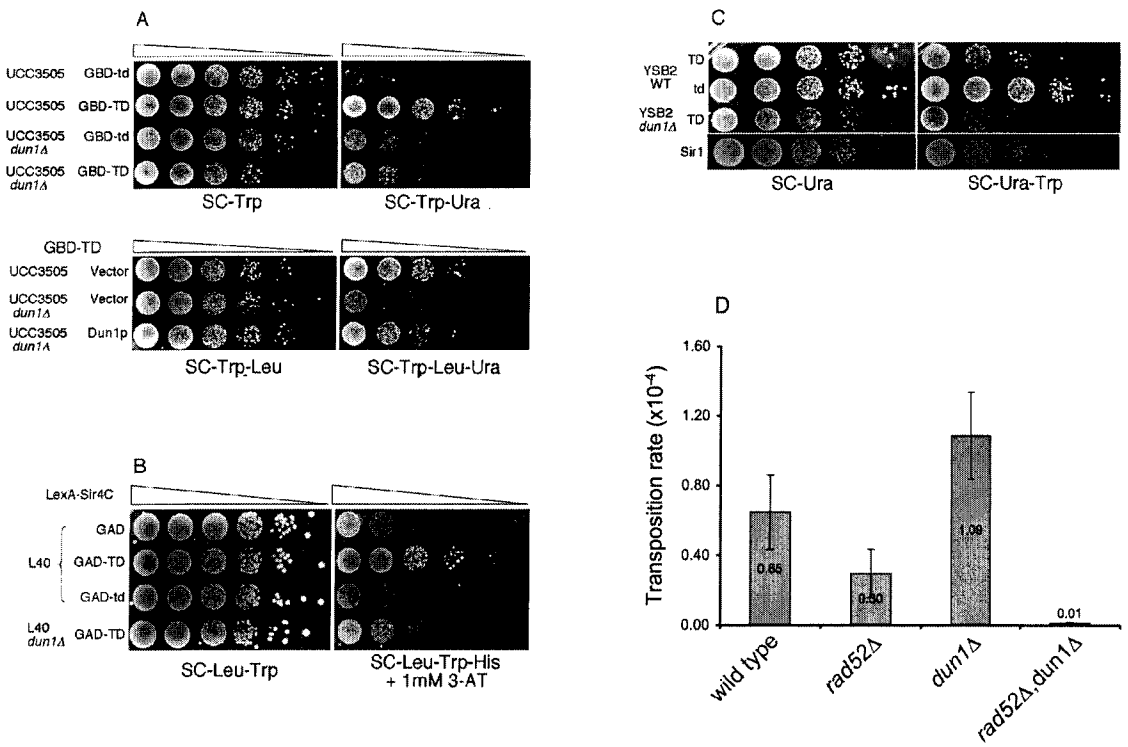

Figure 4. Dun1p affects Ty5 transposition and the interaction between Sir4C and TD. (A) Overexpression of GBD-TD cannot break telomeric silencing in *dun1*Δ, whereas it does break silencing in wild type. An episomal copy of *DUN1* restores the phenotype in *dun1*Δ. (B) Two-hybrid assay to test interaction between Sir4C and TD. The interaction is disrupted in *dun1*Δ. (C) GBD-TD can establish silencing at a crippled *HMR* locus in both wild type and *dun1*Δ. GBD-Sir1p is used as a control. (D) *dun1*Δ increases Ty5 recombination but greatly reduces its transposition. The transposition efficiency is calculated by dividing the number of His^+ colonies by the total number of colonies on YPD plates (Gai and Voytas, 1998).

Dun1p on TD phosphorylation is leaky. To make sure that the loss of telomeric silencing was due to the absence of Dun1p, an episomal copy of the gene was introduced into the knockout strain and the phenotype was successfully restored. Another checkpoint kinase Chk1p (Liu et al., 2000) and the kinase Sat4p partially abrogated two-hybrid interactions (data not shown).

We further tested if loss of Dun1p affects Ty5 transposition. Both *dun1*Δ and *dun1*Δ *rad52*Δ strains were constructed in a YPH499 background. In the wild type strain, the His^+ colonies we obtained are about 0.65×10^{-4}, which includes both the recombination and integration events (Figure 4). By eliminating recombination in the *rad52*Δ strain, Ty5

Table 3. Selected targeting assay results in the strains of kinase knockout.

Strains	pNK254 with wildtype Ty5*					pWW79 with S1095E Ty5*					
4742α#	71/980	57/720	46/588		7.6%		8/90	6/94			7.6%
4742a#	31/421	43/417	19/304		8.1%	18/255	24/317	17/193			7.7%
101E6	1/208		2/403		0.5%		3/343	0/145	2/248		0.7%
103A11	3/252	1/295	3/364		0.8%			4/288	2/223		1.2%
114F11	20/777	17/662	34/1244		2.6%	13/476					2.7%
117B8	3/154	1/116	2/80	2/141	1.6%	3/60	1/81	1/82			2.2%
129D8	6/233	2/92	3/115		2.5%	3/99	2/53	1/49			3.0%
145B9	33/1174	16/692	26/906		2.7%	11/361	8/218	4/191			3.0%
110G12	13/339	13/399	9/360		3.2%	3/151	12/168	18/244	8/137	6/115	5.8%
115F2	9/262	7/232	7/245		3.1%	5/115	10/120	6/70	35/479	6/61	7.3%
117F6	38/1059	12/564	26/686		3.3%	18/289	25/365	13/303			5.9%
119D2	33/685	30/793	10/341		4.0%	25/314	28/458	19/353	12/150	4/196	6.1%
124A9	25/924	44/1254	31/731		3.4%	100/1311	49/684	42/417			7.9%
136F12	30/772	26/838	52/1091		4.0%	75/745	26/209	68/466			11.9%
137D4	37/899	50/1091	41/1081		4.2%	36/531	70/784	11/226			7.6%
138F2	38/782	44/911	56/1101		4.9%	17/233	13/154	32/349			8.4%

These strains are reconstructed by crossing of BY4742 and YDV578. The mating types were marked by **a** and **α**.

* The numbers indicate the number of white colonies/total colonies from individual plates. The percentages were calculated as the sum of white colonies divided by the sum of the totals.

transposition was found to approximate a rate of about $0.3x10^{-4}$, which is consistent with previous studies (Zhu et al., 1999). Remarkably, if we simply delete the *DUN1* kinase, the number of His^+ colonies is doubled. However, this increase in the His^+ colonies is not due to the elevation of Ty5 transposition. Instead, it may be because of an increase in recombination, since in the *dun1*Δ *rad52*Δ double mutant, Ty5 transposition is almost eliminated.

All the data suggest that Dun1p likely phosphorylates TD. To test this, we performed *in vitro* kinase assays with Dun1p and GST-TD purified from *E. coli.* No kinase activity was observed using multiple reaction conditions (data not shown). We also tested the Ty5 C-terminus as a substrate, but Dun1p was unable to phosphorylate this substrate either. We cannot rule out the possibility that factors required for Dun1p to phosphorylate TD are missing in the *in vitro* assay; however, when we used Sml1p – a known Dun1p substrate – which was expressed and purified from *E. coli,* it was phosphorylated very effectively by Dun1p (data not shown).

DISCUSSION

Transposable elements are potent mutagenic agents when active. To avoid the deleterious mutations caused by transposable element insertion, the host may have evolved ways to regulate TE activity. For example, most TEs in a genome are inactive either due to mutation or epigenetic silencing (Slotkin et al., 2005; Yoder et al., 1997). In addition, the integration sites for most TEs are not random, which may be controlled either by the host, the element, or both. For the most part, how integration specificity is controlled is still a mystery. Here we demonstrate that the post-translational modification of Ty5 integrase, which subsequently alters its interaction with Sir4p, is one of the mechanisms that controls integration site choice.

The targeting domain is phosphorylated in vivo

Ty5 encodes a short domain to guide its integration. This domain interacts with Sir4p *in vivo* and *in vitro*. Several lines of genetic and biochemical evidence indicate that TD is phosphorylated. First, the *in vitro* interaction between TD and Sir4C requires that TD is

expressed in yeast cells, and the binding is disrupted by treatment with a protein phosphatase (Figure 1). Furthermore, purification of TD and subsequent MS/MS analysis showed that the second serine within TD is phosphorylated (Figure 2). Finally, replacement of the serine with threonine or glutamic acid partially retained TD function. Importantly, Ty5 integration specificity is not altered when the second serine is replaced by glutamic acid (Table1).

As previously reported (Xie et al., 2001), the interaction between Ty5 integrase and Sir4p is the driving force for targeted integration. Four assays were used to demonstrate this interaction: tethered silencing, loss of silencing, two-hybrid and *in vitro* interaction assays. In the tethered silencing and loss of silencing assays, the binding partner for TD is full-length Sir4p; in the other assays, it is the Sir4p C-terminus (950-1358). In the loss of silencing assays, the output is not the direct result of the interaction. Rather, transcriptional silencing at the telomeres occurs because of degradation of Sir4p, which, in turn, depends on the presence and accumulation of GBD-TD and on Ris1p and other Sir proteins (Fuerst and Voytas, in preparation). These differences may explain the inconsistency of the substitution mutants in the loss of silencing assay relative to the other assays. The threonine and glutamic acid substitutions did not disrupt telomeric silencing, although they tethered silencing to a level comparable to wild type TD. This suggests that the phosphorylated serines might be essential for Sir4p turnover. The S1095E mutant functioned better in tethered silencing and Ty5 targeting than the S1094E mutant, which is consistent with the MS data showing only the second serine is phosphorylated. However, the S1095E mutation could not bind Sir4C in two-hybrid and in *in-vitro* binding assays (Xie and Voytas, data not shown). This may be due to the different forms of Sir4p involved in the interaction (full-length Sir4p vs. Sir4C). It may also be because serine phosphorylation is different from the glutamic acid substitution.

Phosphorylation is required for Ty5 target specificity

The role of phosphorylation at S1095 was demonstrated by SPR spectroscopy in which only the peptide with phosphorylation at this site was able to interact with Sir4C. The unmodified peptide and the peptide with phosphorylation at the first serine could not bind to Sir4C (Figure 4). However, the interaction between the phosphorylated peptide and Sir4C is very weak. With the concentration of peptide as high as 2mM, we still could not saturate the

binding. The dissociation rate is also very high, and after washing, only a very small amount of peptide was still bound to Sir4C, which again, indicates the low binding affinity between the peptide and Sir4C. We speculate that the peptides, because of their short length, do not interact with Sir4p in the same manner as full-length IN. This can be tested in the future by using longer peptides in the SPR experiments.

The original targeting mutant was a S1094L substitution (Gai and Voytas, 1998), and MS analysis indicated that the S1094L mutant does not affect phosphorylation of S1095 (Table 2). Therefore, loss-of-function in the S1094L mutant is not due to the lack of phosphorylation. Interestingly, threonine substitution at position 1094 retains partial activity, which may indicate that the hydroxyl group is required, since either leucine, cysteine or alanine substitutions totally abolished TD function. This would suggest that both the hydroxyl-group at position 1094 and phosphorylated S1095 are required to interact with Sir4p. Alternatively, the existence of the hydroxyl-group before the phosphorylated serine residue may be critical for the three dimensional structure of TD, which allows it to interact with Sir4C. Further experiments are required to distinguish among these possibilities.

Besides phosphorylation, glycosylation is another form of post-translational modification that occurs on serine residues. However, despite our careful analysis of the MS data, we did not find any peaks corresponding to the glycosylated peptide. Therefore, we believe TD is not glycosylated although we cannot rule out the possibility that glycosylated peptides were missed during purification in which only the total soluble proteins were used.

Kinases regulate Ty5 targeting

We used a plasmid-based targeting assay to identify kinases that directly or indirectly affect targeted integration. The assay quantitatively measures the proportion of Ty5 integration events that occur on an episomal *HMR* locus. We expected that mutations in factors important for silencing would result in targeting defects; however, there are no reports of kinases affecting silencing except Mec1p, which is an essential gene and not included in the collection analyzed (Craven and Petes, 2000). Surprisingly, 28 of the 97 kinases tested (about one third) significantly altered targeting. We therefore undertook a second screen,

taking advantage of the S1095E mutant, which mimicks phosphorylation by the negatively charged amino acid and which also targets like the wild type element. Thus, mutations that affect S1095 phosphorylation should be complemented by the S1095E mutant. Eight candidate kinases were identified from the secondary screen, among which three regulate silencing and the others have varied effects on the TD/Sir4p interaction in vivo. Among these, Dun1p significantly decreased the loss of telomeric silencing caused by GBD-TD expression and abrogated the TD/Sir4C two-hybrid interaction. However, we believe that Dun1p is redundant and/or indirectly acts on TD, because loss of Dun1p only moderately reduced Ty5 targeting (to ~5%) and in a *dun1*D strain, TD was still able to tether silencing. Further experiments are needed to determine whether and how the candidate kinases modify TD or affect the integrity of heterochromatin.

Many of the kinases identified are involved in responses to environmental stimuli. Three of the eight candidate kinases are involved in stress responses -- *sat4*Δ is sensitive to 1M NaCl (Mulet et al., 1999), Irs1p (Miyahara et al., 1998) is involved in the PKC pathway, and *mkk1*Δ (Irie et al., 1993) is sensitive to glycerol and nitrogen starvation. Dun1p and Chk1p are checkpoint proteins, and Tpk3p (Garrett and Broach, 1989) is a subunit of the cAMP-dependent protein kinase complex. These latter kinases respond to internal signals. The functions of the kinases imply that Ty5 integration is regulated by the environment. It is possible that in certain circumstances, Ty5 in silent chromatin is expressed, resulting in an "explosion" (sudden and notable multiplication) of the element that causes dramatic changes to the host genome. Therefore, the evolution of the organism will be accelerated to adapt to the new environment.

The role of Dun1p and DNA damage checkpoint kinases in Ty5 targeting and transposition

It has been known for a long time that DNA damage can activate transcription and transposition of the yeast Ty1 retrotransposons (Staleva Staleva and Venkov, 2001). Here we provide the first evidence that Ty5 targeting and transposition is tied to the DNA damage response. Specifically, we demonstrate that the checkpoint kinase, Dun1p, is required for TD and Sir4C interaction, as measured by the yeast two-hybrid and loss of telomeric silencing assays. However, loss of Dun1p does not interfere with the ability of GBD-TD to nucleate

silencing at a crippled *HMR*-E silencer (Figure 1C). This inconsistency may be assay-specific or because of leaky regulation of Dun1p. We also demonstrate that Ty5 transposition is greatly reduced in a *dun1*D strain. Interestingly, the frequency of recombination increases when Dun1p is absent. In the presence of DNA damage, Dun1p is activated by Mec1p/Rad53p (ATM/Chk2 in humans), which phosphorylate both Crt1p, a cofactor for the repression of DNA damage genes, and Sml1p, a repressor of ribonucleotide reductase (RNR). Phosphorylation of Crt1p induces DNA damage gene expression. Phosphorylated Sml1p is degraded, which activates RNRs and subsequently increases the cellular dNTP pools to facilitate DNA repair. Increased dNTP pools may be required for DNA excision repair, or more likely, translesion DNA synthesis (TLS) (Chabes et al., 2003). In the absence of Dun1p, the DNA damage repair genes can still be activated via the Mec1p/Chk1p/Pds1p pathway, but RNR repression is not released. Therefore, the dNTP pools remain at relatively low levels. A possible model is that at low dNTP pools, yeast rely more on homologous recombination to repair DNA damage instead of excision repair or TLS. As a consequence, more His^+ colonies are generated by elevated homologous recombination despite the fact that Ty5 transposition is inhibited in a *dun1*Δ background.

How Dun1p regulates Ty5 transposition and targeting is still unclear. In a *dun1*Δ strain the interaction between TD and Sir4C is disrupted and Ty5 transposition is impaired, however, the failure of Dun1p to phosphorylate the integrase C-terminus (including TD) in the *in vitro* kinase assay suggests that Dun1p does not act on integrase directly. In addition, there is no detectable interaction between Dun1p and TD or the integrase C-terminus (data not shown). Preliminary data indicates that Dun1p is able to phosphorylate Sir4C (Dai and Voytas, unpublished data). Therefore, it is possible that Dun1p regulates Ty5 transposition and targeting by modifying Sir4p and affecting the structure of heterochromatin. More experiments are required to test this hypothesis. Like Dun1p, the other DNA damage checkpoint kinases cannot phosphorylate Ty5 integrase C-terminus either. Furthermore, Ty5 transposition cannot be activated by genotoxic reagents such as methyl methanesulfonate (MMS), 4-nitro quinolineoxide (4-NQO) or hydroxyurea (HU) (Gao and Voytas, unpublished data). It seems, therefore, that the activation of Ty5, unlike Ty1 and Ty3, is not induced by signaling through the DNA damage response pathway. The effects of Dun1 on

Ty5 targeting and transposition may be Dun1p-specific and reveal a novel function of Dun1p.

Post-translational modification of viral proteins

Post-translational modifications play important roles in regulating retroelement activity. Myristylation of the Gag proteins of most retroviruses is necessary, although not sufficient, for the binding of Gag to the plasma membrane (Spearman et al., 1997). Acytalytion of retroviruses, as in the case of avian sarcoma-leukosis virus (ASLV), is found at the amino-terminal methionine, although its biological significance has not been accessed (Palmiter et al., 1978). In contrast, N-acetyltransferase action on viral structural proteins is absolutely required for the intracytoplasmic assembly and infectivity of certain yeast viruses (e.g. L-A double-stranded virus) (Tercero et al., 1993; Tercero and Wickner, 1992). Phosphorylation is another common post-translational modification found on several retroviruses. It has been reported that the Gag protein of HIV-1 is phosphorylated, and more specifically, phosphorylation of tyrosine and serine on matrix (MA) blocks the ability to bind membranes. During viral entry, the phosphorylated MA proteins are associated with the integration complex (Bukrinskaya et al., 1996; Gallay et al., 1996; Gallay et al., 1995a; Gallay et al., 1995b). Phosphorylation is proposed to reveal the nuclear localization signal contained within HIV-1 MA protein (Bukrinsky et al., 1993a; Bukrinsky et al., 1993b), and to enable the preintegration complex to enter the nucleus of nondividing cells (von Schwedler et al., 1994). To date, there are no reports of post-translational modifications of viral proteins other than the Gag.

In this paper, we provide evidence that phosphorylation is important in regulating Ty5 target specificity. Besides phosphorylation of TD, we also observed that other residues in Ty5 INC are phosphorylated *in vitro* by multiple kinases (Gao and Voytas, unpublished). Furthermore, the kinases that phosphorylate INC also affect Ty5 transposition and integration specificity. We expect that phosphorylation of Ty5 INC may not only determine Ty5 integration specificity, but also affect Ty5 transposition or regulate other steps in the Ty5 life cycle. Post-translational modification of retroelement proteins is therefore likely to be an important mechanism by which these elements interact with their host.

EXPERIMENTAL PROCEDURES

Plasmids and Strains

The *Saccharomyces cerevisiae* strains used in this study are as follows. The kinase knockout strains were picked up individually from yeast haploid deletion library (Research Genetics Inc.) in BY4742 background. The strain yDV578 used to reconstruct the kinase knockout strains were obtained through plasmid-mediated gene deletion in yPH499 (Ausubel et al.). The strain yDV629 used for all the protein expression was kindly given by D. Amberg, Upstate Medical University.

The plasmids expressing His6-tagged TD fused to GST were constructed by inserting into the BamHI and SalI sites of pEG(KG) (Mitchell et al., 1993) a short DNA fragment created from two complementary oligonucleotides (DVO3182: 5'-GATCCATCGAAGGTCGTTTGGATTCATCGCCTCCAAATACCTCACATCACCATCACCATCACTAA-3' and DVO3183: 5'-TCGATTAGTGATGGTGATGGTGATGTGAGGTATTTGGAGGCGATGAATCCAAACGACCTTCGATG-3' for making C-terminal His6-tagged TD; DVO31845'-GATCCATCGAAGGTCGTTTGGATTTATCGCCTCCAAATACCTCACATCACCATCACCATCACTAA-3' and DVO31855'-TCGATTAGTGATGGTGATGGTGATGTGAGGTATTTGGAGGCGATAAATCCAAACGACCTTCGATG-3' for making His6-tagged td (S1094L); DVO33115'-GATCCATCGAAGGTCGTTTGGATTCAGCGCCTCCAAATACCTCACATCACCATCACCATCACTAA-3' and DVO33125'-TCGATTAGTGATGGTGATGGTGATGTGAGGTATTTGGAGGCGCTGAATCCAAACGACCTTCGATG-3' for making C-terminal His6-tagged td (S1095A); DVO 33135'-GATCCATCGAAGGTCGTCATCACCATCACCATCACTTGGATTCATCGCCTCCAAATACCTCATAA-3' and DVO33145'-TCGATTATGAGGTATTTGGAGGCGATGAATCCAAGTGATGGTGATGGTGATGACGACCTTCGATG-3' for making N-terminal His6-tagged TD).

In vitro binding and in vitro assays

The *in vitro* binding assay is performed as described in our previous paper (Xie 2001). To treat the immuno-precipitated GBD fusion proteins with λPPase or factor Xa, the final

wash was performed with 1 X enzyme buffer instead of 1 X PBS. The beads were incubated at 30°C with 0.5μl (200 units) enzyme for 30 minutes and washed with 1 X PBS before mixing with the labeled Sir4C or before running a PAGE gel.

Targeting assay

The targeting assay was described previously (Gai and Voytas, 1998), except that here we used the target plasmid pXW72 with a *TRP1* marker rather than pXW78 with a *LEU2* marker.

Protein purification

All steps for purification are carried at 4°C. Cells expressing the fusion proteins are broken by glass bead using a bead beaker in the buffer used for GST-kinase purification as describe (Zhu et al., 2000). Proteins are at first purified using HIS-select Nickle Affinity Gel (Sigma) following the supplied protocol under native condition using BIO-RAD BioLogic LP System. Elutions are combined and purified again with Glutathione-agarose (Sigma) following the supplied protocol and eluted in buffer containing 20mM reduced glutathione, 50mM Tris-HCl, pH9.5±0.1 and 20% glycerol. The fusion proteins are incubated at room temperature with 2μl (20 units) factor Xa for six hours. The digestion mixture is diluted in the equilibration buffer: 50mM sodium phosphate, pH8.0, with 0.3 M sodium chloride and without imidazole, and purified again using HIS-select Nickle Affinity Gel (Sigma, trial scale). The elution is subjected to Mass Spectrometry analysis.

Mass spectrometry

MALDI-TOF MS/MS MS analyses were performed using a QSTAR XL quadrupole TOF mass spectrometer (AB/MDS Sciex, Toronto, Canada) equipped with an oMALDI ion source. The mass spectrometer was operated in the positive ion mode. Mass spectra for MS analysis were acquired over m/z 500 to 4000. After every regular MS acquision, MS/MS acquision was performed against most intensive ions. The molecular ions were selected by information dependent acquiring in the quadrupole analyzer and fragmented in the collision cell. All samples were desalted and concentrated with a 10 μl ZipTipc_{18} (Millipore, Bedford, MA), following the instructions provided by the manufacturer. Peptides were eluted with 1 μl

of 50% acetonitrile/0.1% TFA saturated with α-CHCA and deposited onto the MALDI target plate.

BIAcore surface plasmon resonance analysis

Real-time protein-protein interactions were examined using a BIAcore T100 instrument (BIAcore). Anti-GST antibody was immobilized on a CM5 sensor chip using a GST capture kit (BIAcore). GST-Sir4C was purified from *E. coli* (50mM HEPES pH 7.6, 350mM NaCl, 1mM DTT, 20mM Glutathione, 25% Glycerol), diluted to 10 μg/ml in HBS-EP buffer and injected with a constant (10 μl/min) flow rate at 25°C for 20 min. Synthetic TD or phospho-TD (2.5 mM) was injected at a flow rate of 30 μl/min for 60 seconds at 25°C as analytes respectively. Due to the fast off-rate of all peptides tested, the same GST-Sir4C was used to compare affinity of all three peptides with sufficient time between peptides to allow baseline re-establishment. The chemical binding surface is regenerated with regeneration solution (10 mM glycine-HCl, pH2.2) for 2 min at 20 μl/min. Sensorgrams were subjected to global analysis using BIAcore T100 evaluation software 1.0. Two independent binding comparisons were performed and the data averaged.

ACKNOWLEDGEMENTS

We would like to acknowledge William Lewis and Siquan luo at Proteomic Facility, Plant Science Institute, Iowa State University, for carrying out the MS and MS/MS analysis. This work was supported by National Institutes of Health Grant GM061657 (to D.V.).

REFERENCES:

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1987). Current Protocols in Molecular Biology (New York, Greene/Wiley Interscience).

Boeke, J. D., and Devine, S. E. (1998). Yeast retrotransposons: finding a nice quiet neighborhood. Cell *93*, 1087-1089.

Bukrinskaya, A. G., Ghorpade, A., Heinzinger, N. K., Smithgall, T. E., Lewis, R. E., and Stevenson, M. (1996). Phosphorylation-dependent human immunodeficiency virus type 1 infection and nuclear targeting of viral DNA. Proc Natl Acad Sci U S A *93*, 367-371.

Bukrinsky, M., Sharova, N., and Stevenson, M. (1993a). Human immunodeficiency virus type 1 2-LTR circles reside in a nucleoprotein complex which is different from the preintegration complex. J Virol *67*, 6863-6865.

Bukrinsky, M. I., Sharova, N., McDonald, T. L., Pushkarskaya, T., Tarpley, W. G., and Stevenson, M. (1993b). Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency virus type 1 with viral nucleic acids following acute infection. Proc Natl Acad Sci U S A *90*, 6125-6129.

Bushman, F., Lewinski, M., Ciuffi, A., Barr, S., Leipzig, J., Hannenhalli, S., and Hoffmann, C. (2005). Genome-wide analysis of retroviral DNA integration. Nat Rev Microbiol *3*, 848-858.

Bushman, F. D. (2003). Targeting survival: integration site selection by retroviruses and LTR-retrotransposons. Cell *115*, 135-138.

Chabes, A., Georgieva, B., Domkin, V., Zhao, X., Rothstein, R., and Thelander, L. (2003). Survival of DNA damage in yeast directly depends on increased dNTP levels allowed by relaxed feedback inhibition of ribonucleotide reductase. Cell *112*, 391-401.

Chalker, D. L., and Sandmeyer, S. B. (1992). Ty3 integrates within the region of RNA polymerase III transcription initiation. Genes Dev *6*, 117-128.

Check, E. (2002). A tragic setback. Nature *420*, 116-118.

Chien, C. T., Buck, S., Sternglanz, R., and Shore, D. (1993). Targeting of SIR1 protein establishes transcriptional silencing at HM loci and telomeres in yeast. Cell *75*, 531-541.

Craven, R. J., and Petes, T. D. (2000). Involvement of the checkpoint protein Mec1p in silencing of gene expression at telomeres in Saccharomyces cerevisiae. Mol Cell Biol *20*, 2378-2384.

Devine, S. E., and Boeke, J. D. (1996). Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. Genes Dev *10*, 620-633.

Gai, X., and Voytas, D. F. (1998). A single amino acid change in the yeast retrotransposon Ty5 abolishes targeting to silent chromatin. Mol Cell *1*, 1051-1055.

Gallay, P., Stitt, V., Mundy, C., Oettinger, M., and Trono, D. (1996). Role of the karyopherin pathway in human immunodeficiency virus type 1 nuclear import. J Virol *70*, 1027-1032.

Gallay, P., Swingler, S., Aiken, C., and Trono, D. (1995a). HIV-1 infection of nondividing cells: C-terminal tyrosine phosphorylation of the viral matrix protein is a key regulator. Cell *80*, 379-388.

Gallay, P., Swingler, S., Song, J., Bushman, F., and Trono, D. (1995b). HIV nuclear import is governed by the phosphotyrosine-mediated binding of matrix to the core domain of integrase. Cell *83*, 569-576.

Garrett, S., and Broach, J. (1989). Loss of Ras activity in Saccharomyces cerevisiae is suppressed by disruptions of a new kinase gene, YAKI, whose product may act downstream of the cAMP-dependent protein kinase. Genes Dev *3*, 1336-1348.

Hematti, P., Hong, B. K., Ferguson, C., Adler, R., Hanawa, H., Sellers, S., Holt, I. E., Eckfeldt, C. E., Sharma, Y., Schmidt, M., *et al.* (2004). Distinct genomic integration of MLV and SIV vectors in primate hematopoietic stem and progenitor cells. PLoS Biol *2*, e423.

Hunter, E., Casey, J., Hahn, B., Hayami, M., Korber, B., Kurth, R., Neil, J., Rethwilm, A., Sonigo, P., and Stoye, J. (2000). Retroviridae. In Virus Taxonomy: Seventh Report of the International Committee on Taxonomy of Viruses, M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carsten, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle, and R. B. Wickner, eds. (New York, Academic Press), pp. 369-387.

Irie, K., Takase, M., Lee, K. S., Levin, D. E., Araki, H., Matsumoto, K., and Oshima, Y. (1993). MKK1 and MKK2, which encode Saccharomyces cerevisiae mitogen-activated protein kinase-kinase homologs, function in the pathway mediated by protein kinase C. Mol Cell Biol *13*, 3076-3083.

Kim, J. M., Vanguri, S., Boeke, J. D., Gabriel, A., and Voytas, D. F. (1998). Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. Genome Res *8*, 464-478.

Kimmelman, J. (2005). Recent developments in gene transfer: risk and ethics. Bmj *330*, 79-82.

Kirchner, J., Connolly, C. M., and Sandmeyer, S. B. (1995). Requirement of RNA polymerase III transcription factors for in vitro position-specific integration of a retroviruslike element. Science *267*, 1488-1491.

Kootstra, N. A., and Verma, I. M. (2003). Gene therapy with viral vectors. Annu Rev Pharmacol Toxicol *43*, 413-439.

Lesage, P., and Todeschini, A. L. (2005). Happy together: the life and times of Ty retrotransposons and their hosts. Cytogenet Genome Res *110*, 70-90.

Liu, Y., Vidanes, G., Lin, Y. C., Mori, S., and Siede, W. (2000). Characterization of a Saccharomyces cerevisiae homologue of Schizosaccharomyces pombe Chk1 involved in DNA-damage-induced M-phase arrest. Mol Gen Genet *262*, 1132-1146.

Marshall, E. (2001). Gene therapy. Viral vectors still pack surprises. Science *294*, 1640.

Mitchell, D. A., Marshall, T. K., and Deschenes, R. J. (1993). Vectors for the inducible overexpression of glutathione S-transferase fusion proteins in yeast. Yeast *9*, 715-722.

Mitchell, R. S., Beitzel, B. F., Schroder, A. R., Shinn, P., Chen, H., Berry, C. C., Ecker, J. R., and Bushman, F. D. (2004). Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol *2*, E234.

Miyahara, K., Hirata, D., and Miyakawa, T. (1998). Functional interaction of Isr1, a predicted protein kinase, with the Pkc1 pathway in Saccharomyces cerevisiae. Biosci Biotechnol Biochem *62*, 1376-1380.

Mulet, J. M., Leube, M. P., Kron, S. J., Rios, G., Fink, G. R., and Serrano, R. (1999). A novel mechanism of ion homeostasis and salt tolerance in yeast: the Hal4 and Hal5 protein kinases modulate the Trk1-Trk2 potassium transporter. Mol Cell Biol *19*, 3328-3337.

Palmiter, R. D., Gagnon, J., Vogt, V. M., Ripley, S., and Eisenman, R. N. (1978). The NH2-terminal sequence of the avian oncovirus gag precursor polyprotein (Pr76gag). Virology *91*, 423-433.

Sandoval Rodriguez, A. S., Salazar Montes, A. M., and Armendariz-Borunda, J. (2005). [Viral vectors in gene therapy. Advantages of the adenoassociated vectors]. Rev Gastroenterol Mex *70*, 192-202.

Schroder, A. R., Shinn, P., Chen, H., Berry, C., Ecker, J. R., and Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. Cell *110*, 521-529.

Slotkin, R. K., Freeling, M., and Lisch, D. (2005). Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. Nat Genet *37*, 641-644.

Spearman, P., Horton, R., Ratner, L., and Kuli-Zade, I. (1997). Membrane binding of human immunodeficiency virus type 1 matrix protein in vivo supports a conformational myristyl switch mechanism. J Virol *71*, 6582-6592.

Staleva Staleva, L., and Venkov, P. (2001). Activation of Ty transposition by mutagens. Mutat Res *474*, 93-103.

Tercero, J. C., Dinman, J. D., and Wickner, R. B. (1993). Yeast MAK3 N-acetyltransferase recognizes the N-terminal four amino acids of the major coat protein (gag) of the L-A double-stranded RNA virus. J Bacteriol *175*, 3192-3194.

Tercero, J. C., and Wickner, R. B. (1992). MAK3 encodes an N-acetyltransferase whose modification of the L-A gag NH2 terminus is necessary for virus particle assembly. J Biol Chem *267*, 20277-20281.

von Schwedler, U., Kornbluth, R. S., and Trono, D. (1994). The nuclear localization signal of the matrix protein of human immunodeficiency virus type 1 allows the establishment of

infection in macrophages and quiescent T lymphocytes. Proc Natl Acad Sci U S A *91*, 6992-6996.

Winzeler, E. A., Shoemaker, D. D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J. D., Bussey, H., *et al.* (1999). Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science *285*, 901-906.

Wu, X., Li, Y., Crise, B., and Burgess, S. M. (2003). Transcription start regions in the human genome are favored targets for MLV integration. Science *300*, 1749-1751.

Xie, W., Gai, X., Zhu, Y., Zappulla, D. C., Sternglanz, R., and Voytas, D. F. (2001). Targeting of the Yeast Ty5 Retrotransposon to Silent Chromatin Is Mediated by Interactions between Integrase and Sir4p. Mol Cell Biol *21*, 6606-6614.

Yoder, J. A., Walsh, C. P., and Bestor, T. H. (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends Genet *13*, 335-340.

Zhu, H., Klemic, J. F., Chang, S., Bertone, P., Casamayor, A., Klemic, K. G., Smith, D., Gerstein, M., Reed, M. A., and Snyder, M. (2000). Analysis of yeast protein kinases using protein chips. Nat Genet *26*, 283-289.

Zhu, Y., Dai, J., Fuerst, P. G., and Voytas, D. F. (2003). Controlling integration specificity of a yeast retrotransposon. Proc Natl Acad Sci U S A *100*, 5891-5895.

Zhu, Y., Zou, S., Wright, D. A., and Voytas, D. F. (1999). Tagging chromatin with retrotransposons: target specificity of the Saccharomyces Ty5 retrotransposon changes with the chromosomal localization of Sir3p and Sir4p. Genes Dev *13*, 2738-2749.

Zou, S., Ke, N., Kim, J. M., and Voytas, D. F. (1996). The Saccharomyces retrotransposon Ty5 integrates preferentially into regions of silent chromatin at the telomeres and mating loci. Genes Dev *10*, 634-645.

Zou, S., and Voytas, D. F. (1997). Silent chromatin determines target preference of the Saccharomyces retrotransposon Ty5. Proc Natl Acad Sci U S A *94*, 7412-7416.

CHAPTER 4. A SYSTEMATIC ANALYSIS OF KINASES INVOLVED IN SILENCING IN *SACCHAROMYCES CEREVISIAE*

A manuscript to be submitted to *Genetics*

Junbiao Dai[1] and Daniel F. Voytas[2]

ABSTRACT

The 109 non-essential kinases of *Saccharomyces cerevisiae* were investigated for their roles in maintaining/establishing heterochromatic silencing. Each kinase gene was deleted in two different strains: one with a subtelomeric *URA3* marker (UCC3505, a derivative of S288c) and the other with three different markers at telomeres, *HMR* and the rDNA (CCFY101, a derivative of W303a). Expression of marker genes was assayed on selective media. Depending on the strain, different sets of kinases were identified, which when deleted, disrupt silencing at at least one locus, most frequently the telomeres. With the exception of *tel1*Δ, telomere length was not altered in the kinase deletion strains, suggesting that loss of telomeric silencing is not due to shortened telomeres. Interestingly, most kinases identified are members of MAP kinase pathways involved in osmolarity, cell wall integrity and pheromone responses. We speculate that heterochromatic silencing is regulated by protein kinase networks that respond to environmental changes.

INTRODUCTION

Protein phosphorylation regulates many cellular processes, such as progress through the cell cycle and key events in morphogenesis. In eukaryotes, all protein phosphorylation is conducted by a large protein kinase superfamily. The completion of genome sequencing projects made it possible to determine the protein kinase complement (kinome) in a particular organism. Comparison of kinomes of different organisms generates valuable information on the evolution of protein phosphorylation (reviewed by (Manning *et al.*, 2002a)). About a decade ago, the yeast kinome was the first eukaryote kinome determined (Hunter and

---

[1] Primary researcher and author

[2] Professor and corresponding author, Department of Genetics, Development and Cell Biology, Iowa State University, Ames, IA 50011

Plowman, 1997), and it contains 135 possible kinases (Hunter et al., 2000). An analysis of the kinome of human genome revealed 518 putative protein kinase genes, constituting about 1.7% of all human genes (Manning *et al.*, 2002b). While kinomes are relatively easy to define, deciphering the complex network of phosphorylation-based signaling is more important and much harder.

Silent chromatin is defined as a chromatin domain that is not conducive to transcription. In higher organisms, it is also referred as heterochromatin. Genes located within heterochromatin show position-dependent transcriptional silencing, a well-known phenomenon called position effect variegation (PEV) in *Drosophila* (Eissenberg, 1995). In yeast, heterochromatin is not visible cytologically, but three loci are known to form silent chromatin: telomeres, rDNA and the silent mating-type loci (*HMR* and *HML*). Yeast silent chromatin is a DNA-protein complex, comprised of DNA, DNA binding proteins and the silencing information regulators (Sir proteins, Sir1-4) (recently reviewed by (Rusche *et al.*, 2003)). Silent chromatin at the telomeres and *HM* loci are quite similar but fundamentally differ from that at the rDNA. Genes located within all three loci are transcriptionally silenced. Silent chromatin can be disrupted by mutations in genes encoding some of the protein components such as the Sir proteins, and this results in expression of normally silenced genes.

Silent chromatin, however, is not completely inaccessible to proteins such as transcription factors. Telomeric silencing can be disrupted by expression of the transactivator Ppr1p, which leads to expression of a *URA3* gene at telomere VII-L (Aparicio and Gottschling, 1994). This suggests that the cell may have evolved ways to overcome transcriptional silencing within or close to silent chromatin. Seripauperin (PAU) is a family of genes involved in maintaining cell wall and plasma membrane integrity (Heinisch *et al.*, 1999). These genes are predominantly located at subtelomeric regions (Viswanathan *et al.*, 1994), and their transcriptional activation during stress requires the cell to suppress subtelomeric silencing. Recently, hyperphosphorylation of Sir3p has been shown to cause derepression of PAU genes (Ai *et al.*, 2002). This provides evidence that heterochromatic silencing may be regulated by a protein kinase network in response to environmental

changes.

In this paper, we systematically analyze the effect of kinase deletions on maintaining/establishing heterochromatic silencing in budding yeast. Several protein kinases, when deleted, were shown to have weakened heterochromatic silencing to different extents. Telomere shortening was not the major factor underlying the loss of silencing. Furthermore, in different strain backgrounds, different kinases were involved in the regulation of heterochromatic silencing.

MATERIALS AND METHODS

Strains

Strain UCC3505 (*MATa ade2-101 his3-200 leu2-1 lys2-801 trp1-63 ura3-52 ppr1::HIS3 adh4::URA3-TEL-VIIL DIA5-1)* was a kind gift from D. Gottschling. Strain CCFY101 was a kind gift from K Runger. The yeast haploid deletion library was purchased from Research Genetics Inc.

Re-construction of the kinase deletion strains

We designed two primers for each kinase flanking the ORF (primers typically recognized sites about 500bp up- and downstream of the start and stop codon, respectively). Each pair of primers was used to amply the DNA fragment from the corresponding kinase knockout strain. The resulting fragment contained the $G418^r$ gene and flanking sequences. The primer sequences can be found in supplementary materials. The amplified fragment was used to transform UCC3505 or CCFY101, following a standard transformation protocol (Ausubel *et al.*, 1987). After heat shock, the cells were resuspended in YPD media and grown at 30°C for three hours before plating onto YPD media containing G418 (300 μg/ml). The resistant colonies were picked and purified again on a new YPD plate containing G418. The putative knockout colonies were tested by PCR to confirm the knockout of the kinase with another upstream primer and an internal primer specific to the $G418^r$ gene. The confirmed colonies were grown for additional 50 generations before testing. At least two independent knockout clones were tested for each kinase.

Silencing assay

Cells were grown overnight, diluted to OD_{600} 1.0 and spotted onto either selective media or complete media as described (Roy and Runge, 2000; Xie *et al.*, 2001).

Southern hybridization

Yeast genomic DNA was prepared as described, digested with XhoI and used to prepare a Southern filter (Ausubel *et al.*). The probe was made by PCR amplification of Y' and labeled with ^{32}P using the Rediprime II labeling system (Amersham Biosciences). The Y' primers used for PCR amplification were DVO3048: 5'-GCACTAGCTGTGGAGAGAATGCTGC-3' and DVO3049: 5'-GCTCTTTGTGAACCGCTACCA-3'. Southern hybridizations were performed as described (Ausubel *et al.*) with minor changes and analyzed using phosphor imaging (Typhoon 9410 system, Amersham Biosciences).

RESULTS

Kinase knockout strains affect telomeric silencing

We initially screened a collection of kinase knockout strains for those that affect Ty5 integration specificity. This screen used an established targeting assay (Gai and Voytas, 1998) in which integration of Ty5 onto a plasmid with silent chromatin is monitored. We found that some kinases had a strong influence on Ty5 targeting when they were deleted from the genome (Dai et al., manuscript in preparation). To test if mutations in these kinases affect telomeric silencing, we knocked out these kinases in UCC3505, a strain bearing telomeric *URA3* and *ADE2* markers (Singer and Gottschling, 1994). In wild type UCC3505, the *URA3* gene is silenced very efficiently, and no cells are able to grow in media without uracil. Among eight kinases that alter Ty5 target specificity, we found three of them have strong defects on telomeric silencing as measured by growth on media lacking uracil (*ppr2*Δ, *mkk1*Δ and *tpk3*Δ, data not shown). This result may explain why in these strains Ty5 targeting efficiency was reduced: silent chromatin was probably impaired in these mutants and therefore Ty5 could not recognize silent regions anymore.

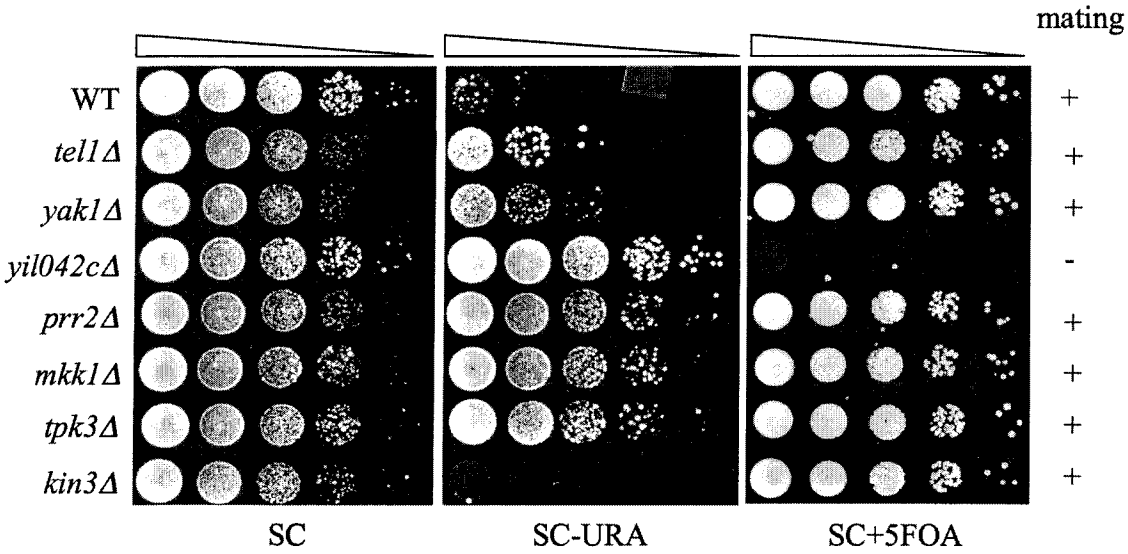

Figure 1. Telomeric silencing is tested in a strain containing the *URA3* reporter gene at the left telomere of chromosome VII. The expression of *URA3* is monitored both on medium lacking uracil (SC-Ura) and on medium containing 5-fluoro-orotic acid (5FOA). Serial, 10-fold dilutions of cells are plated onto the control media and different selective media. *kin3*Δ is used as a control, in which telomeric silencing is still intact. +, mates; -, cannot mate.

The telomeric silencing defect in the kinase deletion mutants suggests that these kinases are involved in the regulation of heterochromatin either directly or indirectly. They may affect silencing by direct modification of heterochromatin components such as histones and other DNA binding proteins; they may also affect silencing by modifying some transcription factors or chromatin-remodeling factors, thereby suppressing heterochromatic silencing. Besides the three kinases we identified above, we speculated that there might be more kinases involved in regulating heterochromatic silencing. Therefore, we systematically knocked out each kinase in UCC3505 and asked whether telomeric silencing was impaired.

In the yeast genome, there are a total of 135 putative kinases (Hunter and Plowman, 1997). Among them, 109 are viable and present in the yeast haploid deletion library (Research Genetics Inc.). To take advantage of the telomeric markers in UCC3505, we reconstructed 100 kinase deletion strains in this strain using a modified PCR method. For the other nine kinases, no resistant colonies could be generated, despite performing multiple

transformations. Some of these kinases are known to be slow-growing and others are temperature sensitive. Of the 100 confirmed deletion strains, we spotted them individually onto selective media to test for telomeric silencing (Figure 1). Six of the kinases were found to have effects on telomeric silencing to different extents, including *ppr2*Δ, *mkk1*Δ and *tpk3*Δ, which were identified in our initial screen. These kinases are discussed individually below. We also tested the ability of these strains to mate, which requires silencing of the *HML* and *HMR* loci. We found that only one strain, *ydl042c*Δ, lost its mating ability.

TEL1: One of the kinase mutants that suppressed telomeric silencing, *tel1*Δ, has been well-studied over the past two decades (reviewed by (Zakian, 1996)). Tel1p is involved in specific DNA damage checkpoints during S-phase as well as in telomere homeostasis (Harrington, 2005). Mutations in *TEL1* result in shortened telomeres (Lustig and Petes, 1986), with minor effects on telomere position effect (TPE) (Boulton and Jackson, 1998). Our result on anti-silencing in *tel1*Δ is also quite weak, consistent with this report. The shortened telomeres may result in the weakened telomeric silencing in *tel1*Δ, although the real reason for the observed TPE is yet unknown.

YAK1: Loss of Yak1p also had only a minor effect on telomeric silencing. *YAK1* is a DYRK (dual specificity Yak1-related kinase) family kinase and is a part of the glucose-sensing system involved in growth control in response to glucose availability (Moriya et al., 2001). In response to glucose signaling, Yak1p translocates from the cytoplasm to the nucleus, where it phosphorylates Pop2p (Moriya et al., 2001), a gene required for expression of glucose-repressed genes in *S. cerevisiae* (Sakai et al., 1992). In addition, Yak1p is negatively regulated by the TOR (target of rapamycin)-PKA (protein kinase A) pathway. In the absence of active Yak1p, Kfh1p binds Fhl1p at ribosomal protein (RP) gene promoters and activates transcription (Martin et al., 2004; Schmelzle et al., 2004). Therefore, Yak1p appears to be an important regulator of RP gene expression.

YIL042C is an uncharacterized protein kinase that is predicted to be located in mitochondria. In a recent survey of deletion mutants that affect telomere length, *yil042c*Δ

was shown to have a slightly shortened telomere (Askree et al., 2004). This is consistent with what we found when we analyzed telomere lengths in the mutants (see below).

PRR2 was identified as a pheromone-response regulator (PRR) in a study that overexpressed 120 protein kinases and tested for perturbations in the pheromone response (Burchett et al., 2001). Interestingly, in the same screen, YAK1 was also identified as a PRR with a slightly weaker phenotype compared to PRR2. It is possible that these two kinases act similarly in regulating telomeric silencing.

MKK1 is a MEK (MAP kinase kinase) in yeast, redundant with *MKK2*. *MKK1/MKK2*, along with upstream *PKC1*, *BCK1* (MEKK) and downstream *SLT2* (MAPK), controls a highly-conserved cell wall integrity signaling pathway (Lee et al., 1993; Martin et al., 1996; Sussman et al., 2004). The MAPK downstream of MKK1, *SLT2*, is known to phosphorylate Sir3p (Ai et al., 2002; Ray et al., 2003). Absence of *SLT2* causes reduced telomeric silencing and enhanced *HMR* and rDNA silencing. Although *MKK1* and *MKK2* are thought to be redundant in the pathway, the silencing defect seems to be specific to *MKK1*, because deletion of *MKK2* has no effect on telomeric silencing (data not shown).

TPK3 encodes one of the subunits of cytoplasmic cAMP-dependent protein kinase (also known as PKA), which contains redundant catalytic subunits Tpk1p, Tpk2p and Tpk3p and the regulatory subunit Bcy1p (Toda et al., 1987a; Toda et al., 1987b). Although the three PKAs share high levels of amino acid sequence identity and have overlapping roles in viability, they are not functionally redundant (Robertson et al., 2000). The only known function of Tpk3p is to inhibit pseudohyphal differentiation, whereas Tpk1p and Tpk2p all have functions in addition to regulating pseudohyphal differentiation (Pan and Heitman, 1999; Robertson et al., 2000). How Tpk3p regulates telomeric silencing is unclear, but Tpk1p has been shown to coordinate with Yak1p to control ribosomal protein gene expression (Martin *et al.*, 2004; Schmelzle *et al.*, 2004). Tpk3p may share a similar function with Tpk1p and regulate telomeric silencing by a mechanism that involves Yak1p.

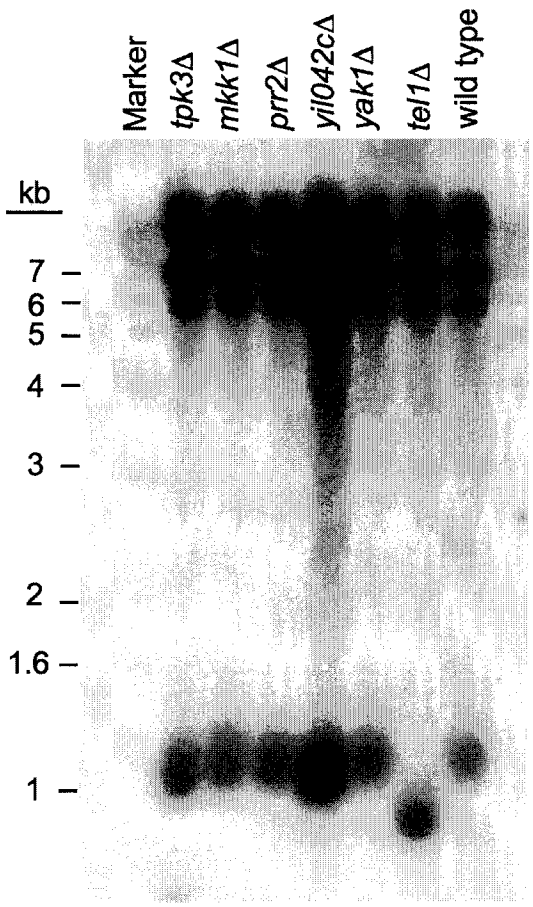

Figure 2. Southern blot of genomic DNA digested with *Xho*I and electrophoresed on a 0.8% agarose gel. The blot was hybridized with a Y′-element probe that corresponds to a subtelomeric repeat present at most yeast telomeres. The sizes of molecular weight markers are indicated on the left of the blot.

Telomere length is not altered in kinase knockout strains with telomeric silencing defects

One reason for the observed telomeric silencing defects may be shortening of telomere length. For example, Tel1p is important to maintain telomere length, and in a *tel1*Δ mutant, the telomere is shortened and telomeric silencing is slightly affected (Boulton and Jackson, 1998). It is possible that other kinases also cause telomere shortening that subsequently leads to loss of telomeric silencing.

The telomere length in each of the kinase knockout strains was measured by Southern hybridization analysis with a probe that recognizes the Y' subtelomeric repeats (Figure 2).

We found the telomere length in *tel1*Δ mutant is significantly shorter than wild type, which is consistent with previous reports (Lustig and Petes, 1986). Beside *tel1*Δ, only *tpk3*Δ and *yil042*Δ mutants showed a slight decrease in telomere length, suggesting that the derepression of telomeric silencing is not caused primarily by telomere shortening. The slightly shortened telomere we observed in the *yil042c*Δ strain is consistent with what Askree et al. found in a recent study of genes affecting telomere length (Askree et al., 2004). Our results indicate that the loss of telomeric silencing in these kinase knockout strains is not due primarily to shortened telomeres. Other mechanisms may play more important roles.

Effects of strain background on silencing by kinases

Two kinase genes, *SLT2* and *BCK1*, are known to cause defects in telomeric silencing if deleted (Ray *et al.*, 2003). However, we did not recover these genes in our screen. This raised the possibility that the silencing phenotype is strain specific, since the test strain used in the other study was derived from W303a and ours (CCFY101) was derived from S288C. We therefore carried out another screen using W303a (Ray *et al.*, 2003). This time, we successfully reconstructed 84 kinase deletion strains. The other 25 strains were either not able to produce G418^r colonies (12 kinases) or could not be confirmed by PCR (13 kinases). Among these latter strains was *yak1*Δ. We therefore carried out another screen using W303a (Ray *et al.*, 2003). This time, we successfully reconstructed 84 kinase deletion strains. The other 25 strains were either not able to produce G418^r colonies (12 kinases) or could not be confirmed by PCR (13 kinases). Among these latter strains was *yak1*Δ.

In contrast to the first screen, silencing in the W303a kinase deletion strains was monitored at three different loci, namely the telomeres, *HM*R and rDNA. As expected, most kinase deletions had no effect on silencing; however, 12 kinase deletion strains showed significant perturbation in silencing at at least one locus (Figure 3). These included three kinases (*SLT2*, *BCK1* and *DUN1*) known to impair silencing in W303a if mutated. The identification of these kinases indicated that our screen was effective. Other than *TEL1*, none of the other four kinases that affected telomeric silencing in UCC3505 had an effect in W303a (*YAK1* was not tested).

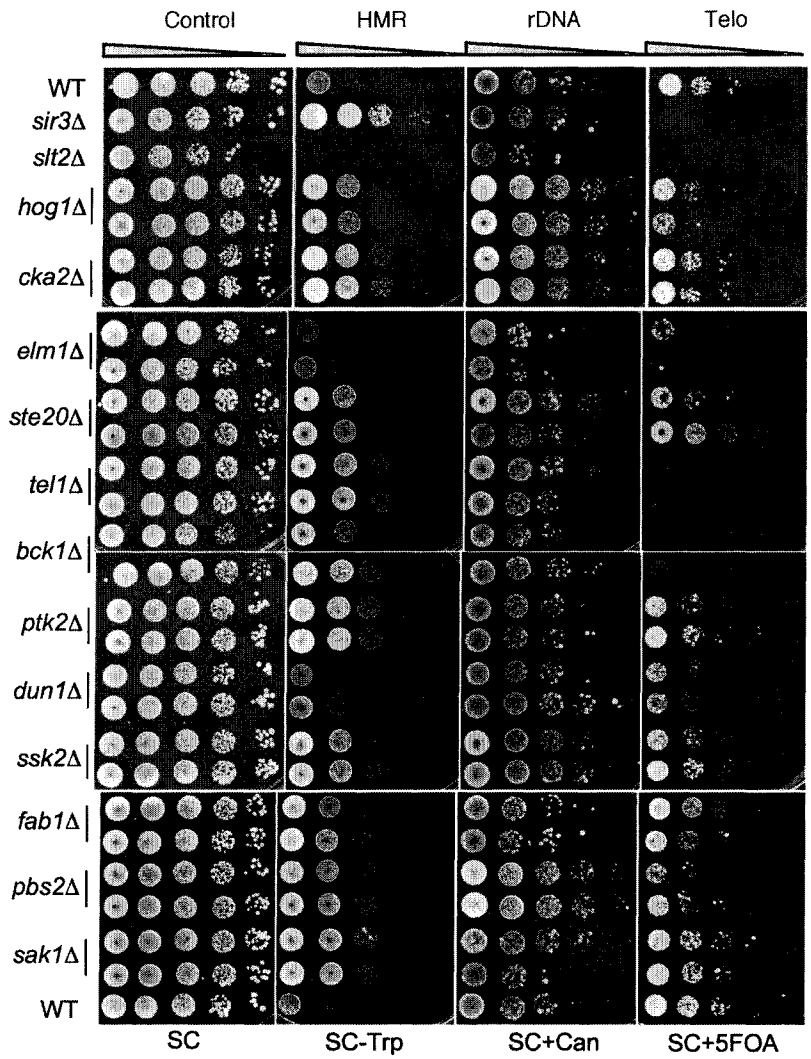

Figure 3. Silencing is tested in the triple silencer strain with three different reporter genes. Serial, 10-fold dilutions of cells were plated onto control (SC) and different selective media.
Expression of *TRP1* in the *hmrE* locus is monitored by growth on medium lacking tryptophan (SC-Trp), where more growth equals less silencing. Expression of *CAN1* in the rDNA locus is monitored using negative selection on media containing canavanine (SC+Can), so that more growth equals more silencing. Expression of *URA3* placed adjacent to the right telomere of chromosome V is monitored by negative selection on media containing 5-FOA (SC+5FOA), so that more growth equals more silencing. Two independent knockout clones are tested for each kinase.

TEL1 is the only kinase identified in both screens. Compared to its phenotype in UCC3505, the telomeric silencing defect was significantly greater in W303a. As discussed above, our observation in UCC3505 is consistent with a previous report (Boulton and Jackson, 1998), in which a strain similar to UCC3505 was used. It is possible that some co-factors required to maintain telomeric silencing are not present in the W303a derivative

CCFY101. Therefore, in the absence of Tel1p, telomeric silencing is greatly reduced. Besides telomeric silencing, *HMR* silencing was also slightly impaired in the *tel1*Δ mutant, suggesting that Tel1p may be involved in maintaining *HMR* silencing too. rDNA silencing was not obviously affected by *TEL1* deletion.

***BCK1* and *SLT2*:** In a previous report, strains bearing *BCK1* or *SLT2* deletions showed defects in telomeric silencing, but increased *HMR* and rDNA silencing (Ray *et al.*, 2003). We also found that telomeric silencing was strongly decreased in both mutants. However, *HMR* and rDNA silencing were not significantly affected, even though the same strain was used. This discrepancy may result from differences in assay conditions. The effect on telomeric silencing in *slt2*Δ was different from another report, in which they claimed that deletion of *SLT2* (*MPK1*) does not significantly affect subtelomeric silencing (Ai *et al.*, 2002). This difference may be due to a strain difference, since in this study, a strain similar to UCC3505 was used. As discussed above, *MKK1* affected telomeric silencing in UCC3505; however, it didn't affect silencing in CCFY101 even though it is upstream of *SLT2* and downstream of *BCK1*. Probably, in CCFY101, *MKK1* shares a redundant function with *MKK2* to regulate heterochromatic silencing.

***SSK2*, *PBS2* and *HOG1*:** These three kinases are known players in the high osmolarity glycerol (HOG) response pathway. Upon osmotic stress, Ssk2p is phosphorylated and activated. Active Ssk2p phosphorylates Pbs2p, which in turn phosphorylates Hog1p. Phosphorylated Hog1p then translocates into the nucleus and activates osmoresponsive genes (reviewed in (Schwartz and Madhani, 2004)). Mutations in either of the three genes causes reduced *HMR* silencing, indicating they may act in the same pathway to regulate *HMR* silencing. Telomeric silencing was weakened in *pbs2*Δ and slightly weakened in *hog1*Δ, but not affected in *ssk2*Δ. All three kinase mutants showed strengthened silencing at the rDNA, but to different extents, suggesting a redistribution of silencing factors in the nucleus when these factors are mutated.

DUN1 is a DNA damage checkpoint kinase, which has been shown to cause defects in telomeric silencing when deleted in a strain derived from W303a (Craven and Petes, 2000;

Mallory *et al.*, 2003). This phenotype was confirmed in our screen. Both telomeric and *HMR* silencing were weakened in *dun1*Δ. Besides Dun1p, another checkpoint protein, Mec1p, is also involved in regulation of telomeric silencing (Craven and Petes, 2000). Since *mec1*Δ is lethal, we did not test it in our screen. rDNA silencing was not affected in *dun1*Δ.

ELM1: A strong effect on telomeric silencing was observed in *elm1*Δ strains. These strains also had slightly weakened rDNA silencing, but no effect was observed on *HMR* silencing. Elm1p regulates cellular morphogenesis, septin behavior and cytokinesis (Bouquin *et al.*, 2000; Garrett, 1997; Koehler and Myers, 1997). Similar to *tpk3*Δ strains, *elm1*Δ strains have constitutive filamentous growth (Edgington *et al.*, 1999).

***CKA2*:** Weakened *HMR* silencing was observed in *cka2*Δ, but no effect was found for telomeric silencing. *cka2*Δ *strains* had strengthened rDNA silencing, similar to what was observed in *hog1*Δ and *pbs2*Δ strains.

***STE20, PTK2, FAB1* and *SAK1*:** All four of these kinases showed minor effects on silencing. They had slightly weakened *HMR* silencing, but showed no effect on rDNA or telomeric silencing.

We also tested if telomere length was altered in the W303a kinase mutants. Similar to what we observed in the UCC3505 background, there was no significant change in telomere length except for *tel1*Δ (data not shown).

DISCUSSION

We identified several kinases, which when deleted, affect heterochromatic silencing. Among these kinases are a few that have previously been shown to affect telomeric silencing in yeast. Mutations on *TEL1*, an ATM ortholog, cause shortened telomeres and slightly weakened telomeric silencing (Boulton and Jackson, 1998). Mutations in the DNA damage checkpoint kinases, Mec1p and Dun1p, also result in derepression of telomeric silencing (Craven and Petes, 2000; Mallory *et al.*, 2003). The underlying mechanisms by which these kinases regulate telomeric silencing are not known. However, our data indicate that telomere

shortening is not the major reason, since in most of these kinase deletion strains, telomere length is not changed (Figure 3).

The recruitment of heterochromatin binding proteins is critical to establish and maintain heterochromatin structures. Therefore, the association/dissociation of such proteins as a consequence of phosphorylation may be one mechanism by which kinases regulate heterochromatic silencing. In addition, phosphorylation of DNA binding proteins such as histones may also interfere with the association/dissociation of heterochromatin binding proteins, causing changes in heterochromatin status. Heterochromatin protein 1 (HP1) – an important component of heterochromatin -- has been shown to be phosphorylated in different organisms (Eissenberg *et al.*, 1994; Huang *et al.*, 1998a). In mammalian cells, phosphorylation of HPgamma might reduce its transcriptional repression activity (Koike *et al.*, 2000). Phosphorylated Drosophila HP1 has been suggested to be required for efficient heterochromatin binding (Zhao and Eissenberg, 1999). Serine-to-alanine and serine-to-glutamate substitutions at consensus protein kinase motifs resulted in reduction or loss of silencing activity of mutant HP1 in transgenic flies (Zhao *et al.*, 2001), suggesting that the dynamic phosphorylation/dephosphorylation regulates HP1 activity. Recently, histone phosphorylation has been shown to block HP1 binding to heterochromatin (Daujat *et al.*, 2005; Fischle *et al.*, 2005; Hirota *et al.*, 2005). Hale et al. also showed that phosphorylation of histone H1 prevents its interaction with HPalpha in mammalian cells (Hale *et al.*, 2006). In yeast, Sir3p is known to be phosphorylated, and phosphorylation of Sir3p results in the redistribution of heterochromatic silencing factors (Ray *et al.*, 2003). The kinases we identified that are involved in regulating heterochromatin may phosphorylate heterochromatin factors. Our preliminary data suggest Sir4p is another target of phosphorylation, but the biological consequence of Sir4p phosphorylation remains to be elucidated (Dai and Voytas, unpublished data).

In addition to components of heterochromatin, transcription factors/DNA binding proteins could be another large group of targets for kinases that regulate heterochromatic silencing. Many transcription factors are phosphorylated. In HeLa cells, the Ca^{2+}-dependent phosphorylation of transcriptional activator Ets-1 reduces its DNA affinity, correlating with

reduced levels of gene transcription (Pufall *et al.*, 2005). Similarly, in response to DNA damage and replication blocks, Crt1p, a DNA-binding protein in yeast, is hyperphosphorylated and loses its ability to bind DNA. The dissociation of Crt1p removes the general repressors Ssn6p and Tup1p, which are bound to the promoters of damage-inducible genes, resulting in transcriptional induction of these normally silenced genes (Huang *et al.*, 1998b). Therefore, by modifying either transcription activators or repressors, kinases are able to regulate the silencing status of particular sets of genes. Some of the kinases we identified may also act to modulate silencing by phosphorylating transcription factors.

Our results also suggest that silencing is regulated differently in different strain backgrounds. This can explain some of the discrepancies observed in past studies. For example, one group reported that mutation of *SLT2* causes significant loss of telomeric silencing in a strain derived from W303a (Ray *et al.*, 2003). However, in an earlier report, it is claimed that deletion of *SLT2* does not result in significant loss of subtelomeric silencing (Ai *et al.*, 2002). The strain used in this latter study was similar to UCC3505, which is derived from S288C. Alternatively, the observed strain differences may be caused by where the marker is located on the chromosome. In UCC3505, the marker is introduced onto the left telomere of chromosome VII, replacing the native telomere with the *URA3* gene followed by $TG_{1\text{-}3}$ repeats (Singer and Gottschling, 1994). In CCFY101, the *URA3* gene and $TG_{1\text{-}3}$ repeats replace the right telomere of chromosome V (Roy and Runge, 2000). Telomeric silencing at different chromosomes may be regulated differently.

Interestingly, most of the kinases we identified are members of MAP kinase pathways. For example, we found three kinases in the high osmolarity growth pathway and three kinases in the cell integrity pathway. Two other kinases (*TPK3* and *YAK1*) are involved in growth control in response to glucose availability. Furthermore, *PAU* genes are involved in maintaining cell wall and plasma membrane integrity (Heinisch *et al.*, 1999), and are predominantly located in subtelomeric regions (Viswanathan *et al.*, 1994). The activation of *PAU* genes during stress requires the cell to suppress subtelomeric silencing. The hyperphosphorylation of Sir3p has been shown to cause derepression of these genes (Ai *et*

al., 2002), and Sir3 is phosphorylated by Slt2p (Ai *et al.*, 2002; Ray *et al.*, 2003), a kinase in the cell integrity pathway. These data lead us to speculate that the derepression of heterochromatic silencing might correspond to the activation of stress genes in response environmental changes that are transduced through kinase-signaling pathways. Understanding of the relationship between stresses, the MAP kinase signaling pathways and heterochromatic silencing will require additional study. The kinases identified here provide a starting point for such future investigations.

ACKNOWLEDGEMENTS

We thank Dr. D. Gottschling and Dr. K. Runger for providing strains. This work was supported by National Institutes of Health Grant GM061657 (to D.V.).

REFERENCES

Ai W, Bertram PG, Tsang CK, Chan TF, Zheng XF. 2002. Regulation of subtelomeric silencing during stress response. Mol Cell 10:1295-1305.

Aparicio OM, Gottschling DE. 1994. Overcoming telomeric silencing: a trans-activator competes to establish gene expression in a cell cycle-dependent way. Genes Dev 8:1133-1146.

Askree SH, Yehuda T, Smolikov S, Gurevich R, Hawk J, Coker C, Krauskopf A, Kupiec M, McEachern MJ. 2004. A genome-wide screen for Saccharomyces cerevisiae deletion mutants that affect telomere length. Proc Natl Acad Sci U S A 101:8658-8663.

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K. 1987. Current Protocols in Molecular Biology. New York: Greene/Wiley Interscience.

Boulton SJ, Jackson SP. 1998. Components of the Ku-dependent non-homologous end-joining pathway are involved in telomeric length maintenance and telomeric silencing. Embo J 17:1819-1828.

Bouquin N, Barral Y, Courbeyrette R, Blondel M, Snyder M, Mann C. 2000. Regulation of cytokinesis by the Elm1 protein kinase in Saccharomyces cerevisiae. J Cell Sci 113 (Pt 8):1435-1445.

Burchett SA, Scott A, Errede B, Dohlman HG. 2001. Identification of novel pheromone-response regulators through systematic overexpression of 120 protein kinases in yeast. J Biol Chem 276:26472-26478.

Craven RJ, Petes TD. 2000. Involvement of the checkpoint protein Mec1p in silencing of gene expression at telomeres in Saccharomyces cerevisiae. Mol Cell Biol 20:2378-2384.

Daujat S, Zeissler U, Waldmann T, Happel N, Schneider R. 2005. HP1 binds specifically to Lys26-methylated histone H1.4, whereas simultaneous Ser27 phosphorylation blocks HP1 binding. J Biol Chem 280:38090-38095.

Edgington NP, Blacketer MJ, Bierwagen TA, Myers AM. 1999. Control of Saccharomyces cerevisiae filamentous growth by cyclin-dependent kinase Cdc28. Mol Cell Biol 19:1369-1380.

Eissenberg JC, Elgin, S. C. R., and Paro, R. 1995. Epigenetic regulation in Drosophila: a conspiracy of silence. In: Elgin SCR, editor. Chromatin Structure and Gene Expression: Oxford University Press. p 141-171.

Eissenberg JC, Ge YW, Hartnett T. 1994. Increased phosphorylation of HP1, a heterochromatin-associated protein of Drosophila, is correlated with heterochromatin assembly. J Biol Chem 269:21315-21321.

Fischle W, Tseng BS, Dormann HL, Ueberheide BM, Garcia BA, Shabanowitz J, Hunt DF, Funabiki H, Allis CD. 2005. Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. Nature 438:1116-1122.

Gai X, Voytas DF. 1998. A single amino acid change in the yeast retrotransposon Ty5 abolishes targeting to silent chromatin. Mol Cell 1:1051-1055.

Garrett JM. 1997. The control of morphogenesis in Saccharomyces cerevisiae by Elm1 kinase is responsive to RAS/cAMP pathway activity and tryptophan availability. Mol Microbiol 26:809-820.

Hale TK, Contreras A, Morrison AJ, Herrera RE. 2006. Phosphorylation of the linker histone H1 by CDK regulates its binding to HP1alpha. Mol Cell 22:693-699.

Harrington L. 2005. Making the most of a little: dosage effects in eukaryotic telomere length maintenance. Chromosome Res 13:493-504.

Heinisch JJ, Lorberg A, Schmitz HP, Jacoby JJ. 1999. The protein kinase C-mediated MAP kinase pathway involved in the maintenance of cellular integrity in Saccharomyces cerevisiae. Mol Microbiol 32:671-680.

Hirota T, Lipp JJ, Toh BH, Peters JM. 2005. Histone H3 serine 10 phosphorylation by Aurora B causes HP1 dissociation from heterochromatin. Nature 438:1176-1180.

Huang DW, Fanti L, Pak DT, Botchan MR, Pimpinelli S, Kellum R. 1998a. Distinct cytoplasmic and nuclear fractions of Drosophila heterochromatin protein 1: their

phosphorylation levels and associations with origin recognition complex proteins. J Cell Biol 142:307-318.

Huang M, Zhou Z, Elledge SJ. 1998b. The DNA replication and damage checkpoint pathways induce transcription by inhibition of the Crt1 repressor. Cell 94:595-605.

Hunter E, Casey J, Hahn B, Hayami M, Korber B, Kurth R, Neil J, Rethwilm A, Sonigo P, Stoye J. 2000. Retroviridae. In: van Regenmortel MHV, Fauquet CM, Bishop DHL, Carsten EB, Estes MK, Lemon SM, Maniloff J, Mayo MA, McGeoch DJ, Pringle CR, Wickner RB, editors. Virus Taxonomy: Seventh Report of the International Committee on Taxonomy of Viruses. New York: Academic Press. p 369-387.

Hunter T, Plowman GD. 1997. The protein kinases of budding yeast: six score and more. Trends Biochem Sci 22:18-22.

Koehler CM, Myers AM. 1997. Serine-threonine protein kinase activity of Elm1p, a regulator of morphologic differentiation in Saccharomyces cerevisiae. FEBS Lett 408:109-114.

Koike N, Maita H, Taira T, Ariga H, Iguchi-Ariga SM. 2000. Identification of heterochromatin protein 1 (HP1) as a phosphorylation target by Pim-1 kinase and the effect of phosphorylation on the transcriptional repression function of HP1(1). FEBS Lett 467:17-21.

Lee KS, Irie K, Gotoh Y, Watanabe Y, Araki H, Nishida E, Matsumoto K, Levin DE. 1993. A yeast mitogen-activated protein kinase homolog (Mpk1p) mediates signalling by protein kinase C. Mol Cell Biol 13:3067-3075.

Lustig AJ, Petes TD. 1986. Identification of yeast mutants with altered telomere structure. Proc Natl Acad Sci U S A 83:1398-1402.

Mallory JC, Bashkirov VI, Trujillo KM, Solinger JA, Dominska M, Sung P, Heyer WD, Petes TD. 2003. Amino acid changes in Xrs2p, Dun1p, and Rfa2p that remove the preferred targets of the ATM family of protein kinases do not affect DNA repair or telomere length in Saccharomyces cerevisiae. DNA Repair (Amst) 2:1041-1064.

Manning G, Plowman GD, Hunter T, Sudarsanam S. 2002a. Evolution of protein kinase signaling from yeast to man. Trends Biochem Sci 27:514-520.

Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. 2002b. The protein kinase complement of the human genome. Science 298:1912-1934.

Martin DE, Soulard A, Hall MN. 2004. TOR regulates ribosomal protein gene expression via PKA and the Forkhead transcription factor FHL1. Cell 119:969-979.

Martin H, Castellanos MC, Cenamor R, Sanchez M, Molina M, Nombela C. 1996. Molecular and functional characterization of a mutant allele of the mitogen-activated protein-

kinase gene SLT2(MPK1) rescued from yeast autolytic mutants. Curr Genet 29:516-522.

Moriya H, Shimizu-Yoshida Y, Omori A, Iwashita S, Katoh M, Sakai A. 2001. Yak1p, a DYRK family kinase, translocates to the nucleus and phosphorylates yeast Pop2p in response to a glucose signal. Genes Dev 15:1217-1228.

Pan X, Heitman J. 1999. Cyclic AMP-dependent protein kinase regulates pseudohyphal differentiation in Saccharomyces cerevisiae. Mol Cell Biol 19:4874-4887.

Pufall MA, Lee GM, Nelson ML, Kang HS, Velyvis A, Kay LE, McIntosh LP, Graves BJ. 2005. Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region. Science 309:142-145.

Ray A, Hector RE, Roy N, Song JH, Berkner KL, Runge KW. 2003. Sir3p phosphorylation by the Slt2p pathway effects redistribution of silencing function and shortened lifespan. Nat Genet 33:522-526.

Robertson LS, Causton HC, Young RA, Fink GR. 2000. The yeast A kinases differentially regulate iron uptake and respiratory function. Proc Natl Acad Sci U S A 97:5984-5988.

Roy N, Runge KW. 2000. Two paralogs involved in transcriptional silencing that antagonistically control yeast life span. Curr Biol 10:111-114.

Rusche LN, Kirchmaier AL, Rine J. 2003. The establishment, inheritance, and function of silenced chromatin in Saccharomyces cerevisiae. Annu Rev Biochem 72:481-516.

Sakai A, Chibazakura T, Shimizu Y, Hishinuma F. 1992. Molecular analysis of POP2 gene, a gene required for glucose-derepression of gene expression in Saccharomyces cerevisiae. Nucleic Acids Res 20:6227-6233.

Schmelzle T, Beck T, Martin DE, Hall MN. 2004. Activation of the RAS/cyclic AMP pathway suppresses a TOR deficiency in yeast. Mol Cell Biol 24:338-351.

Schwartz MA, Madhani HD. 2004. Principles of MAP kinase signaling specificity in Saccharomyces cerevisiae. Annu Rev Genet 38:725-748.

Singer MS, Gottschling DE. 1994. TLC1: template RNA component of Saccharomyces cerevisiae telomerase. Science 266:404-409.

Sussman A, Huss K, Chio LC, Heidler S, Shaw M, Ma D, Zhu G, Campbell RM, Park TS, Kulanthaivel P, Scott JE, Carpenter JW, Strege MA, Belvo MD, Swartling JR, Fischl A, Yeh WK, Shih C, Ye XS. 2004. Discovery of cercosporamide, a known antifungal natural product, as a selective Pkc1 kinase inhibitor through high-throughput screening. Eukaryot Cell 3:932-943.

Toda T, Cameron S, Sass P, Zoller M, Scott JD, McMullen B, Hurwitz M, Krebs EG, Wigler M. 1987a. Cloning and characterization of BCY1, a locus encoding a regulatory subunit of the cyclic AMP-dependent protein kinase in Saccharomyces cerevisiae. Mol Cell Biol 7:1371-1377.

Toda T, Cameron S, Sass P, Zoller M, Wigler M. 1987b. Three different genes in S. cerevisiae encode the catalytic subunits of the cAMP-dependent protein kinase. Cell 50:277-287.

Viswanathan M, Muthukumar G, Cong YS, Lenard J. 1994. Seripauperins of Saccharomyces cerevisiae: a new multigene family encoding serine-poor relatives of serine-rich proteins. Gene 148:149-153.

Xie W, Gai X, Zhu Y, Zappulla DC, Sternglanz R, Voytas DF. 2001. Targeting of the yeast Ty5 retrotransposon to silent chromatin is mediated by interactions between integrase and Sir4p. Mol Cell Biol 21:6606-6614.

Zakian VA. 1996. Structure, function, and replication of Saccharomyces cerevisiae telomeres. Annu Rev Genet 30:141-172.

Zhao T, Eissenberg JC. 1999. Phosphorylation of heterochromatin protein 1 by casein kinase II is required for efficient heterochromatin binding in Drosophila. J Biol Chem 274:15095-15100.

Zhao T, Heyduk T, Eissenberg JC. 2001. Phosphorylation site mutations in heterochromatin protein 1 (HP1) reduce or eliminate silencing activity. J Biol Chem 276:9512-9518.

CHAPTER 5. TRANSPOSITION AND INTEGRATION SPECIFICITY OF TY5 RETROTRANSPOSONS ARE REGULATED BY MULTIPLE PROTEIN KINASES IN *SACCHAROMYCES CEREVISIAE*

Junbiao Dai[*,1], Jiquan Gao[*,2] and Daniel F. Voytas[3]

ABSTRACT

The targeting domain (TD) of Ty5 integrase is phosphorylated, and phosphorylation is required for interaction with Sir4p and targeting of Ty5 insertions to domains of heterochromatin. To identify kinases that phosphorylate Ty5 integrase, a portion of the integrase C-terminus (mINC) that encompasses TD was used as a substrate in *in vitro* kinase reactions. *All* of the kinases encoded by the yeast genome (125 kinases total) were purified as GST-fusion proteins and tested for their ability to phosphorylate mINC. Sixteen kinases were identified that modify mINC, and deletions in genes encoding these kinases affect Ty5 integration specificity. Interestingly, mutations in two kinases (*IME2* and *RCK2*) increased Ty5 integration specificity significantly. The remainder impaired integration specificity as measured by a plasmid-based targeting assay. Using several different mINC constructs with various serine/threonine mutations, the kinases were grouped based on their ability to phosphorylate a defined set of serine/threonine residues. Four kinases – Hrr25p, Rim11p, Rck2p and Yak1p – are the most likely candidates for phosphorylating TD. We also identified two kinases, Yck1p and Yck2p, both of which phosphorylated a serine adjacent to TD (S1091). Phosphorylation of mINC by multiple kinases, and importantly, the observation that mutations in these kinases affect Ty5 integration specificity suggests that both transposition and integration of Ty5 are tightly regulated in response to cellular processes or environmental stress.

---

* These two authors contributed equally to this work
[1] Primary researcher and author
[2] Iowa State University graduate student who participated in most of the experiments
[3] Professor and corresponding author, Department of Genetics, Development and Cell Biology, Iowa State University, Ames, IA 50011

INTRODUCTION

Integration of reverse transcribed cDNA into the host genome is one of the essential steps in the retroelement (including retroviruses and retrotransposons) life cycle. The protein that carries out integration is the element-encoded integrase (IN). IN is characterized by three distinct domains: an N-terminal zinc finger motif (HHCC motif), a catalytic core (DD35E motif) and a C-terminal domain (INC) (Haren et al., 1999). Whereas the zinc finger and central catalytic core are highly conserved among retroelements, the INCs tend to be divergent. During integration, IN nicks both cDNA ends to generate two free 3'-OH, which are used to attack and link to the target DNA on different strands. The resulting gap is repaired by the host DNA repair machinery, leaving a target site duplication flanking the element, which is a hallmark of retrotransposition.

Ty5 IN is about 80 kD and is mostly insoluble (Irwin and Voytas, 2001). IN is released by protease cleavage from a large, precursor polyprotein, however, the precise boundaries of Ty5 IN have not yet been mapped. Nonetheless, based on sequence comparisons with other retroelements, Ty5 IN can be roughly divided into the three domains mentioned above (Peterson-Burch and Voytas, 2002). Ty5 IN contains a comparatively large and extended C-terminus, enriched in proline, asparagine, serine-threonine and aspartate-glutamate. One conserved domain (GKGY motif) within Ty5 INC has been identified (Peterson-Burch and Voytas, 2002). Recent analysis suggests this motif is important for RT function (Brady and Voytas, manuscript in preparation). Besides the GKGY motif, a putative nuclear localization signal (NLS) was also found in Ty5 INC (Fuerst and Voytas, unpublished). Similarly, Ty1 also contains an NLS in this region (Kenna et al., 1998; Moore et al., 1998).

Previous work in our lab has found a determinant of Ty5 integration specificity within INC. This motif, named the targeting domain (TD), is comprised of six amino acids (LDSSPP) and is located close to the end of the INC (Gai and Voytas, 1998; Xie et al., 2001). TD interacts with the Sir4p C-terminus (Sir4C), and this interaction directs the pre-integration complex to heterochromatic regions, giving rise to Ty5's insertion preference (Xie et al., 2001; Zhu et al., 2003). Interestingly, the interaction with Sir4C is strengthened

when larger regions of INC are tested that encompass TD, suggesting that other parts of INC facilitate the binding to Sir4C (Xie et al., 2001). In addition, we recently found that TD is phosphorylated *in vivo* and that phosphorylation is required for Ty5 targeting (Dai et al. manuscript in preparation; Chapter 3). All these data indicate that INC is important for Ty5 targeted integration. Other important but yet undiscovered functions may also be carried out by INC.

In this paper, we purified a complete set of yeast protein kinases and tested if they are able to phosphorylate INC. we found a region of Ty5 INC (mINC) that is phosphorylated by multiple kinases *in vitro*. The biological relevance of this phosphorylation is not totally clear yet; however, when the kinases that modify mINC were deleted individually from the genome, Ty5 transposition and/or integration specificity was altered. This data suggests that *in vivo,* these kinases are also important in regulating Ty5 activity. Using substrates bearing different mutations, we divided these kinases into several groups, which phosphorylate different sites within mINC.

RESULTS AND DISCUSSION

A Portion of INC is phosphorylated by multiple kinases in vitro

We have previously shown that Ty5 TD is phosphorylated *in vivo* and that phosphorylation is required for interaction with Sir4C (Dai et al., manuscript in preparation; Chapter 3). To seek the kinases responsible for TD phosphorylation, we constructed a smaller version of the IN C-terminus (mINC), which spans aa1065-1131. mINC includes TD and flanking amino acid residues. The mINC sequence was modified to encode a six histidine tag and inserted into an *E. coli.* expression vector (see Materials and Methods). The peptide was expressed, purified and used as a substrate in *in vitro* kinase assays. 125 kinases were assayed that represent the complete complement of kinases encoded by the yeast genome. These kinases were purified from yeast as GST-kinase fusion proteins (the fusion library was a kind gift from M. Snyder). Purification was monitored by Western blots using an antibody that recognizes GST. Among the purified kinases, 20 did not show a signal on Western blots. This is consistent with the report describing this kinase collection, in which

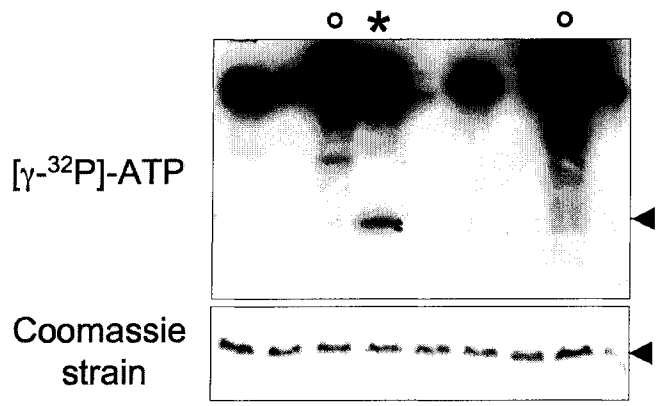

Figure 1. GST-kinases were purified from yeast and then subjected to an *in vitro* kinase assay with mINC as a substrate. The reaction was carried out at 30°C for one hour. The reaction mixtures were separated by PAGE and detected by phosphorimaging. The arrow indicates the substrate position. The star shows a positive reaction. o indicates a potential kinase that requires re-testing. The lower panel shows a coomassie stain of the mINC substrate.

these same 14 kinases also could not be detected by Western analysis (Zhu et al., 2000). However, as shown in this report, even without detectable signals, some of the kinases still displayed high activity. Based on this, we performed *in vitro* kinase assays with all of the purified kinases, including those that gave no signals on Western blots.

A typical result of the *in vitro* kinase assays is shown in Figure 1. The lower panel shows a Coomassie stain of the mINC substrate, and the upper panel is an autoradiogram of the substrates after treatment with the kinase. The star indicates a kinase that phosphorylates the substrate. The circle indicates a candidate kinase that may act on mINC; these kinases were retested to determine if the activity was reproducible. Most of the purified kinases were active, as indicated by the presence of an upper band. This band represents either autophosphorylation of the kinase or phosphorylation of GST. Some kinases, even with large amounts of enzyme or long reaction times (up to three hours) displayed no activity. Three reasons can explain this observation: First, the kinase may not be functional, perhaps due to lack of activation (such as phosphorylation by an upstream kinase); second, the kinase may require co-factors to be active and those co-factors were missing in the reaction; third,

the kinases may simply not phosphorylate mINC, GST or themselves. We cannot distinguish among these possibilities. However, based on the presence of the large band in most reactions and success from previous studies (Zhu *et al.*, 2000), we believe lack of kinase activity is not a major obstacle in our screen for kinases responsible for mINC phosphorylation.

Before performing the *in vitro* kinase assays, we used publicly-accessible bioinformatics tools to predict candidate kinases. Using NetPhos 2.0 (http://www.cbs.dtu.dk/services/NetPhos/) with default parameters, we found nine predicted phosphorylation sites in mINC, including five serines, three threonines and one tyrosine. Interestingly, the second serine within TD, which we identified as a phosphorylation site, was predicted to be phosphorylated at a very high score (0.987). Furthermore, when we used NetPhosK 1.0 (http://www.cbs.dtu.dk/services/NetPhosK/) to predict which kinases are able to phosphorylate mINC, six different kinases families were predicted including PKC (*PKC1*), GSK3 (*MCK1*, *RIM11*, *MRK1*, *YGK3*), CKI (*YCK1*, *YCK2*, *YCK3*, *HRR25*), cdc2 (*CDC2*), cdk5 and p38MAPK (the yeast homologs are shown in parentheses). We therefore paid particular attention to these kinases in the *in vitro* kinase assays. Furthermore, PPSP (http://bioinformatics.lcd-ustc.org/PPSP/) and GPS (http://973-proteinweb.ustc.edu.cn/gps/gps_web/) predicted that GSK3, CKI and MAPKKK family kinases could phosphorylate the second serine (S1095) within TD. However, S1095 is not the only site in mINC predicted to be acted upon by these kinases, suggesting that if we mutate S1095 in mINC, we might not be able to abolish mINC phosphorylation.

Among the 125 kinases tested, 16 were able to phosphorylate mINC (Table 1). Interestingly, and consistent with our predictions, all four yeast casein kinase I (CKI) isoforms phosphorylated mINC. In addition, one GSK3 homolog, Rim11p, was also a mINC kinase. Two catalytic subunits of protein kinase A (Tpk1p and Tpk2p) were also identified. The other interesting kinase was Ste20p, which is the upstream kinase in the pheromone response pathway. Since Ty5 is activated by the pheromone response pathway in the presence of α-factor (Ke *et al.*, 1997), we speculate that some of the kinases in the pheromone response pathway may be involved in regulating Ty5 transposition and/or

Table 1. Kinases involved in phosphorylation of mINC.

Systematic Name	Gene Name	Targeting Efficiency (%)	pJB182*	pJB199*	pJB220*	pJB221*	pJB222*
Wild type		~7.98±2.10					
YHR135C	YCK1	2.63±0.094	+	–	–	–	–
YNL154C	YCK2	5.59±2.2	+	+/-	–	–	–
YER123W	YCK3	n.t.	+	n.t.	–	–	–
YPL204W	HRR25	n.t.	+	+/-	–	+	+
YMR139W	RIM11	6.38±2.91	+	+/-	–	+	+
YJL164C	TPK1	3.36±0.74	+	+/ -	+	+	+
YPL203W	TPK2	3.35±0.67	+	n.t.	+	+	+
YLR096W	KIN2	5.34±0.72	+	n.t.	–	–	–
YAR018C	KIN3	0.49±0.01	+	n.t.	–	–	–
YIL095W	PRK1	2.45±0.24	+	n.t.	–	–	–
YHL007C	STE20	n.t.	+	n.t.	–	–	–
YJL106W	IME2	14.55±5.78	+	n.t.	–	–	–
YLR248W	RCK2	12.59±1.37	+	n.t.	–	+	+
YJL141C	YAK1	6.70	+	n.t.	–	+	+
YPL026C	SKS1	5.45	+	n.t.	–	n.t.	n.t.
YKL139W	CTK1	n.t.	+	n.t.	–	n.t.	n.t.

* pJB182 is the wild type mINC. All other pJBs contain mutations at some positions (Figure 3). n.t., not tested; +, signal; +/–,weak signal; –,no signal

targeting. The other two interesting kinases are Ime2p and Rck2p, which will be discussed in more detail below. Yak1p and Ssk1p are both involved in glucose sensing (Moriya *et al.*, 2001; Vagnoli and Bisson, 1998). Ctk1p is the catalytic subunit of C-terminal domain kinase I (CTDK-I), a cyclin-dependent kinase (Sterner *et al.*, 1995). The other kinases are Prk1p, Kin2p and Kin3p. Although all of these kinases were able to phosphorylate mINC *in vitro*, it does not mean they also phosphorylate Ty5 IN *in vivo*. The biological relevance of these kinases in regulating Ty5 activity still needs more experimental testing.

Kinases that phosphorylate mINC affect Ty5 target specificity

One way to test the biological relevance of the kinases we identified *in vitro* is to test whether Ty5 activity, i.e. transposition and/or targeting, is altered in strains with mutations in these kinases. As previously reported (Chapter 3), we used the kinase deletion strains to screen for kinases that affect TD modification. Here we correlate this data with data from the in vitro kinase assays.

Among the 16 kinases that phosphorylate mINC, we were unable to obtain a strain with a deletion in *STE20*, because its loss results in very slow growth. Similarly, mutations in *HRR25* are lethal, and *hrr25*Δ is not included in the haploid deletion library (Research Genetics Inc). For the other 14 kinases, we found that most significantly affect Ty5 integration (Table 1 and Figure 2). In two kinase deletion strains, *ime2*Δ and *rck2*Δ, Ty5 targeting efficiency increased, whereas for all other strains, targeting was decreased. Ime2p is a protein kinase involved in meiosis that activates the meiosis-specific transcription factor Ndt80p (Sopko *et al.*, 2002). Ime2p expression is activated during sporulation (~33 fold) (Chu *et al.*, 1998), but it is even more dramatically activated when cells are treat with α-factor (~100 fold) (Roberts *et al.*, 2000). Since exposure to α-factor also activates Ty5 transposition, this suggests a link between Ime2p activation and IN phosphorylation. Further experiments will be required to test this linkage.

Rck2p is a protein kinase involved in the response to oxidative and osmotic stress (Bilsland *et al.*, 2004). In contrast to Ime2p, which promotes meiosis, Rck2p is a meiosis inhibitor (Ramne *et al.*, 2000). The expression of Rck2p is inhibited slightly during sporulation

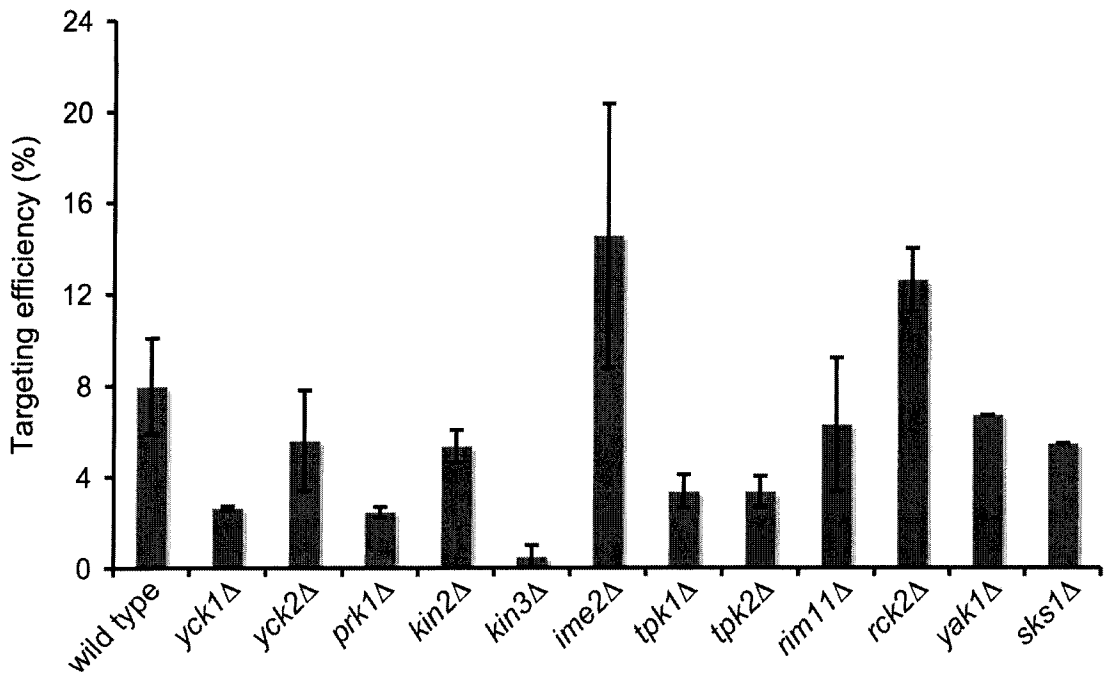

Figure 2. Ty5 targeting efficiency in kinase knockout strains.

(2.7 fold) (Chu *et al.*, 1998), and α-factor treatment has no effect (Roberts *et al.*, 2000). Interesting, Rck2p is not the only member of the oxidative and osmotic stress pathway that affects Ty5 targeting. Loss of Rck1p and Hog1p both result in increased Ty5 targeting efficiency to levels comparable to a *rck2Δ* mutant. Rck1p is a homolog to Rck2p, and Hog1p is a MAP kinase that interacts with Rck2p (Ho *et al.*, 2002; Uetz *et al.*, 2000). Although to date there is no evidence that Ty5 can be activated by environmental stresses, transposition of Ty1 is activated by environmental stresses such as DNA damage, nitrogen starvation (which activates the invasive-filamentous growth pathway, recently reviewed by (Lesage and Todeschini, 2005)) and severe adenine starvation (Todeschini *et al.*, 2005). We speculate that Ty5 may also be activated via oxidative and osmotic stress pathways and that Rck1p, Rck2p and Hog1p play a role in regulating Ty5 activity.

In one kinase deletion strain, *kin3*Δ, Ty5 integration specificity was dramatically decreased from ~8% to <0.5%, suggesting that Kin3p plays an important role in regulating Ty5 integration specificity. Alternatively, this kinase may cause re-distribution of Sir4p, the chromosomal determinant of Ty5 targeting (Xie *et al.*, 2001; Zhu *et al.*, 2003). However, as discussed in Chapter 3, the loss of targeting in *kin3*Δ may be an artifact of the plasmid-based targeting assay. Eight chromosomal Ty5 insertions were recovered from the *kin3*Δ strain, and all were located in preferred Ty5 targets: five in telomeric regions and three at the *HM* loci (data not shown). Therefore, it will be very interesting to determine where the Kin3p phosphorylation sites reside within mINC and the biological function of Kin3p phosphorylation.

In four other kinase deletion strains, *yck1*Δ, *prk1*Δ, *tpk1*Δ and *tpk2*Δ, Ty5 integration specificity was decreased significantly. As discussed above, Yck1p is one of the casein kinase I isoforms in yeast and predicted to be one of the kinases that phosphorylate TD. Besides TD, there are several other sites within mINC that are predicted to be phosphorylated by Yck1p. The decrease in Ty5 integration specificity suggests that Yck1p may indeed play a role in regulating Ty5 activity *in vivo*. The involvement of two TPKs in regulating Ty5 integration specificity indicates that Ty5 targeting may be regulated by nutrition stress and/or filamentous growth. Interestingly, Ty1 activation is controlled by Fus3p and Kss1p, two kinases involved in filamentous growth (Conte *et al.*, 1998; Conte and Curcio, 2000; Morillon *et al.*, 2000). We also observed that in *fus3*Δ and *kss1*Δ strains, Ty5 transposition is activated (Xie and Voytas, unpublished data). Although Fus3p or Kss1p did not phosphorylate mINC *in vitro*, they may regulate Ty5 transposition either by phosphorylating another portion of INC or by regulating other kinases. It will be interesting to determine how the filamentous growth pathway regulates Ty5 transposition and targeting. Finally, it is difficult to speculate how Prk1p, a kinase involved in regulating the organization and function of the actin cytoskeleton (Zeng and Cai, 1999), regulates Ty5 target specificity.

The observed differences in Ty5 integration specificity in various kinase deletion strains suggests that phosphorylation may not simply be involved in directing Ty5 integration to heterochromatin. Certainly, loss of some kinases interferes with TD phosphorylation and

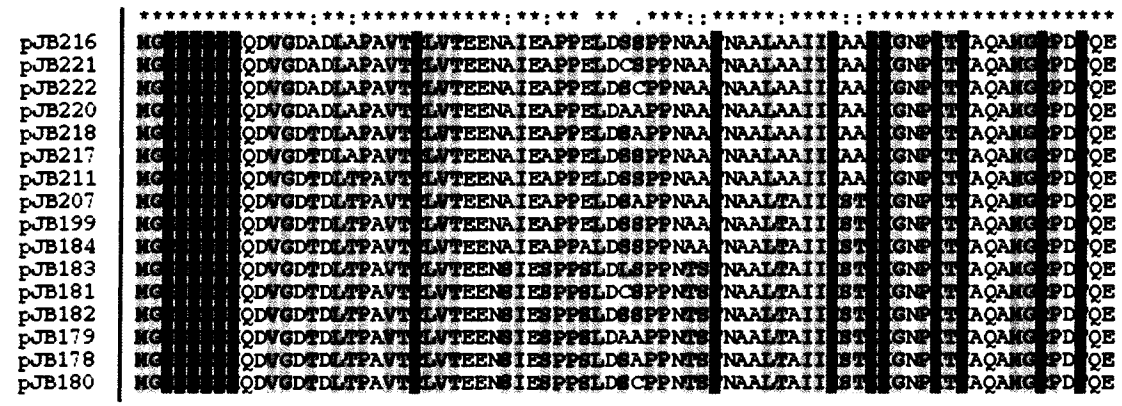

Figure 3. Amino acid sequence alignment of mINC and its mutants

subsequently decreases heterochromatic targeting. However, phosphorylation at other sites may target Ty5 to other chromosomal loci in response to environmental stress. These two possibilities are not mutually exclusive and may collectively regulate Ty5 integration specificity under different conditions.

Mapping the phosphorylation sites within mINC

After we identified the kinases that phosphorylate mINC, we wanted to test whether they specifically phosphorylate the second serine (S1095) within TD. To do this, two mutant versions of mINC were constructed in which serine 1094 (the first serine within TD) and serine 1095 were substituted with cysteine. The mINC variants were expressed and purified from *E. coli* and used as substrates in the *in vitro* kinase assay. Unfortunately, all of the kinases phosphorylated the modified substrates (data not shown). This suggests that either we did not identify the TD kinase or, more likely, that the TD kinase can phosphorylate both a serine within TD and some other serine/threonine residue. In mINC, there are 13 serines/threonines, not including the two serines within TD (Figure 3). Some of these serine/threonine residues were predicted to be phosphorylated at high probability. Therefore, we generated a series of constructs with mutations in various serine and threonine residues throughout mINC (Figure 3). Several of these constructs were used in the tests described below.

One mINC variant (pJB199 with S1985A, S1989A, S1999A, S1991E and T2000A substitutions) was found to eliminate phosphorylation by Yck1p, indicating that these four resides include all possible Yck1p phosphorylation sites. The same construct was also found to lower the phosphorylation signals for the other four CKI kinases (Yck2p, Hrr25p, Rim11p and Tpk1p), suggesting that these sites may also be targets for these other kinases. However, the rest of mINC still contains sites that were phosphorylated by these kinases. Because of this, threonines 1070, 1073, 2012 and serine 2011 in pJB199 were substituted by alanines to generate pJB216 (Figure 3). At the same time, we also mutated S1094 or S1095 to cysteine or mutated both serines to alanine, generating pJB221, pJB222 and pJB220, respectively. In these constructs, only two threonines (T1077 and T1081) remain, and these threonines were not predicted to be phosphorylated (with scores of 0.178 and 0.127, respectively). These three constructs (pJB220, pJB221 and pJB222) were individually tested in assays with 14 of the kinases (all except the two most recently identified kinases, namely Sks1p and Ctk1p).

Based on our results with substrates pJB220, pJB221 and pJB222, we divided the 14 kinases into three groups. The largest group could not phosphorylate these three substrates and includes eight kinases (Yck1p, Yck2p, Yck3p, Kin2p, Kin3p, Prk1p, Ste20p and Ime2p). This result suggests that the phosphorylation sites for these kinases no longer exist on the substrate and further, that these eight kinases do not modify TD. The second group contains two kinases, Tpk1p and Tpk2p, which phosphorylated all three substrates, including the S1094A S2095A mutant. This suggests that the two TPKs do not phosphorylate TD. The last group included four kinases (Hrr25p, Rim11p, Rck2p and Yak1p) that do not phosphorylate the substrate bearing two alanine substitutions within TD (pJB220); however, these kinases were able to modify the other two substrates. This group of kinases was the most interesting since the data suggest these kinases phosphorylate the serines within TD. Given the possibility that when the preferred serine was mutated, the kinases can phosphorylate the serine close to its recognition site, we speculate that at least one of these kinases phosphorylates the second serine within TD.

We performed another experiment to identify phosphorylation sites for some of the kinases using the synthetic peptide (PPSLDSSPPNTS) as a substrate. This peptide was

initially used for *in vitro* binding experiments to test for direct interactions between TD peptides and Sir4C (see Chapter 3 for details). After the *in vitro* kinase reaction, the products were analyzed by mass spectrometry to determine if the peptide was phosphorylated. Of six kinases tested (Yck2p, Rim11p, Tpk1p, Yck1p, Ime2p and Hrr25p), two kinases (Yck1p and Yck2p) phosphorylated the peptide. Further tandem MS analysis showed that both modified the same serine adjacent to TD (S1091, shown in bold in the sequence above). Although we did get a signal indicating the peptide was doubly phosphorylated, the signal was too weak to make any reliable conclusions.

In summary, using the different strategies described above, we were able to narrow down the kinases that phosphorylate TD to four candidates – Hrr25p, Rim11p, Rck2p and Yak1p. Further, we were able to show that the 16 kinases that phosphorylate mINC affect Ty5 integration specificity. Future experiments will include additional MS/MS analysis to identify the specific residues that are modified by these kinases. We anticipate that these studies will identify the kinase that phosphorylates S1095 within TD and thereby determines where in the genome Ty5 integrates. This will open up new lines of investigation to determine how the cell regulates integration specificity, which has important consequences for genome integrity and evolution.

MATERIALS AND METHODS

Cell culture, constructs and protein purification

A DNA fragment encoding mINC was amplified by the polymerase chain reaction (PCR) (Ausubel *et al.*, 1987) from plasmid pNK254, which contains the Ty5 retrotransposon (Ke and Voytas, 1997). The amplification product was cloned into the NcoI/XhoI sites of the expression vector pET28-b (Invitrogen). PCR mutagenesis was used to substitute serines/threonines by alanines at various sites of the mINC (Ausubel *et al.*, 1987). The mINC and various mutants were expressed in *E.coli* strain BL21 RIL (Stratagen) and purified using HIS-select Nickel Affinity Gel (Sigma) following the supplied protocol for purification under native conditions. The GST-kinases were expressed in a *pep4* yeast strain, which lacks several proteases (Mitchell *et al.*, 1993) and purified following the protocol provided by M. Snyder at Yale University.

In vitro kinase assay

The *in vitro* kinase reactions were carried out following the published protocol for a casein kinase activity assay (Ausubel *et al.*, 1987). In brief, 5 μl of kinases and 2 μl of substrate were used for a typical 20 μl reaction. After incubating at 30 °C for one hour, 4 μl of 6x SDS sample buffer was added into each sample and incubated at 95 °C for five minutes. The reaction mixtures were separated by SDS-PAGE, washed and exposed to X-ray film or phosphoimaged.

Ty5 targeting assay

The targeting assay was performed as described previously (Gai and Voytas, 1998).

ACKNOWLEDGEMENTS

We thank Dr. M. Snyder at Yale University for providing us with the complete GST-kinase fusion library and the protocol for purification.

REFERENCES

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1987). Current Protocols in Molecular Biology (New York, Greene/Wiley Interscience).

Bilsland, E., Molin, C., Swaminathan, S., Ramne, A., and Sunnerhagen, P. (2004). Rck1 and Rck2 MAPKAP kinases and the HOG pathway are required for oxidative stress resistance. Mol Microbiol *53*, 1743-1756.

Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P. O., and Herskowitz, I. (1998). The transcriptional program of sporulation in budding yeast. Science *282*, 699-705.

Conte, D., Jr., Barber, E., Banerjee, M., Garfinkel, D. J., and Curcio, M. J. (1998). Posttranslational regulation of Ty1 retrotransposition by mitogen-activated protein kinase Fus3. Mol Cell Biol *18*, 2502-2513.

Conte, D., Jr., and Curcio, M. J. (2000). Fus3 controls Ty1 transpositional dormancy through the invasive growth MAPK pathway. Mol Microbiol *35*, 415-427.

Gai, X., and Voytas, D. F. (1998). A single amino acid change in the yeast retrotransposon Ty5 abolishes targeting to silent chromatin. Mol Cell *1*, 1051-1055.

Haren, L., Ton-Hoang, B., and Chandler, M. (1999). Integrating DNA: transposases and retroviral integrases. Annu Rev Microbiol *53*, 245-281.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., *et al.* (2002). Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature *415*, 180-183.

Irwin, P. A., and Voytas, D. F. (2001). Expression and processing of proteins encoded by the Saccharomyces retrotransposon Ty5. J Virol *75*, 1790-1797.

Ke, N., Irwin, P. A., and Voytas, D. F. (1997). The pheromone response pathway activates transcription of Ty5 retrotransposons located within silent chromatin of Saccharomyces cerevisiae. Embo J *16*, 6272-6280.

Ke, N., and Voytas, D. F. (1997). High frequency cDNA recombination of the *Saccharomyces* retrotransposon Ty5: The LTR mediates formation of tandem elements. Genetics *147*, 545-556.

Kenna, M. A., Brachmann, C. B., Devine, S. E., and Boeke, J. D. (1998). Invading the yeast nucleus: a nuclear localization signal at the C terminus of Ty1 integrase is required for transposition in vivo. Mol Cell Biol *18*, 1115-1124.

Lesage, P., and Todeschini, A. L. (2005). Happy together: the life and times of Ty retrotransposons and their hosts. Cytogenet Genome Res *110*, 70-90.

Mitchell, D. A., Marshall, T. K., and Deschenes, R. J. (1993). Vectors for the inducible overexpression of glutathione S-transferase fusion proteins in yeast. Yeast *9*, 715-722.

Moore, S. P., Rinckel, L. A., and Garfinkel, D. J. (1998). A Ty1 integrase nuclear localization signal required for retrotransposition. Mol Cell Biol *18*, 1105-1114.

Morillon, A., Springer, M., and Lesage, P. (2000). Activation of the Kss1 invasive-filamentous growth pathway induces Ty1 transcription and retrotransposition in Saccharomyces cerevisiae. Mol Cell Biol *20*, 5766-5776.

Moriya, H., Shimizu-Yoshida, Y., Omori, A., Iwashita, S., Katoh, M., and Sakai, A. (2001). Yak1p, a DYRK family kinase, translocates to the nucleus and phosphorylates yeast Pop2p in response to a glucose signal. Genes Dev *15*, 1217-1228.

Peterson-Burch, B. D., and Voytas, D. F. (2002). Genes of the Pseudoviridae (Ty1/copia retrotransposons). Mol Biol Evol *19*, 1832-1845.

Ramne, A., Bilsland-Marchesan, E., Erickson, S., and Sunnerhagen, P. (2000). The protein kinases Rck1 and Rck2 inhibit meiosis in budding yeast. Mol Gen Genet *263*, 253-261.

Roberts, C. J., Nelson, B., Marton, M. J., Stoughton, R., Meyer, M. R., Bennett, H. A., He, Y. D., Dai, H., Walker, W. L., Hughes, T. R., *et al.* (2000). Signaling and circuitry of

multiple MAPK pathways revealed by a matrix of global gene expression profiles. Science *287*, 873-880.

Sopko, R., Raithatha, S., and Stuart, D. (2002). Phosphorylation and maximal activity of Saccharomyces cerevisiae meiosis-specific transcription factor Ndt80 is dependent on Ime2. Mol Cell Biol *22*, 7024-7040.

Sterner, D. E., Lee, J. M., Hardin, S. E., and Greenleaf, A. L. (1995). The yeast carboxyl-terminal repeat domain kinase CTDK-I is a divergent cyclin-cyclin-dependent kinase complex. Mol Cell Biol *15*, 5716-5724.

Todeschini, A. L., Morillon, A., Springer, M., and Lesage, P. (2005). Severe adenine starvation activates Ty1 transcription and retrotransposition in Saccharomyces cerevisiae. Mol Cell Biol *25*, 7459-7472.

Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., *et al.* (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature *403*, 623-627.

Vagnoli, P., and Bisson, L. F. (1998). The SKS1 gene of Saccharomyces cerevisiae is required for long-term adaptation of snf3 null strains to low glucose. Yeast *14*, 359-369.

Xie, W., Gai, X., Zhu, Y., Zappulla, D. C., Sternglanz, R., and Voytas, D. F. (2001). Targeting of the yeast Ty5 retrotransposon to silent chromatin is mediated by interactions between integrase and Sir4p. Mol Cell Biol *21*, 6606-6614.

Zeng, G., and Cai, M. (1999). Regulation of the actin cytoskeleton organization in yeast by a novel serine/threonine kinase Prk1p. J Cell Biol *144*, 71-82.

Zhu, H., Klemic, J. F., Chang, S., Bertone, P., Casamayor, A., Klemic, K. G., Smith, D., Gerstein, M., Reed, M. A., and Snyder, M. (2000). Analysis of yeast protein kinases using protein chips. Nat Genet *26*, 283-289.

Zhu, Y., Dai, J., Fuerst, P. G., and Voytas, D. F. (2003). Controlling integration specificity of a yeast retrotransposon. Proc Natl Acad Sci U S A *100*, 5891-5895.

CHAPTER 6. GENERAL CONCLUSIONS

GENERAL DISCUSSION

Transposable elements (both transposons and retroelements) must integrate a copy of their DNA into the host genome to survive from generation to generation. Random integration, however, is potentially hazardous and could have deleterious genetic effects to the host by acting as a mutagen. Therefore, elements and their host have coevolved mechanisms to regulate retroelement integration. The successful proliferation of retrotransposons is thought to be due to their ability to identify "safe havens" in the genome, which are often gene-poor regions where integration is not harmful to their hosts (Boeke and Devine, 1998; Craig, 1997). Integration into safe-havens can be achieved by three different mechanisms: an active selection of targets by the elements; the forced integration into restricted regions by the host; or a coordinated process regulated by both the element and host. The data presented in this dissertation support the third mechanism, that is, at least in the case of Ty5, integration specificity is regulated cooperatively by both element and its host.

Ty5 integration specificity is determined by the integrase targeting

Integration of retroelement cDNA into the host genome is catalyzed by an element-encoded enzyme called integrase (IN). IN is comprised of three distinct domains: an N-terminal zinc finger (HHCC motif), a catalytic core (DD35E motif) and a C-terminal domain (INC) (Haren et al., 1999). Among LTR retroelement INs, the zinc finger motif and central catalytic core region are highly conserved; however, the C-terminal domain is very divergent. This suggests the C-terminus is evolving very rapidly, and therefore it would be predicted to carry out element-specific functions that act independently from the conserved, catalytic functions of IN. In agreement with this hypothesis, the comparatively large and extended C-terminus of Ty5 IN contains a GKGY motif (Peterson-Burch and Voytas, 2002), a putative nuclear localization signal (NLS) (Fuerst and Voytas, unpublished data) and a targeting domain (TD) (Xie et al., 2001). Interactions between TD and the heterochromatin protein Sir4p target Ty5 integration to domains of heterochromatin (Zou et al., 1996) (Xie et al., 2001).

Several lines of evidence suggest TD functions as an independent module. First, the interaction between IN and Sir4p is primarily dependent on TD. Even a point mutation (S1094L) within TD abolishes the interaction with Sir4p and greatly reduces Ty5 targeting (Gai and Voytas, 1998). In addition, new Ty5 integration hotspots are created when Sir4p is tethered to ectopic DNA sites. Targeting to sites of tethered Sir4p is abrogated by single amino acid substitutions in either TD or Sir4p that prevent their interaction (Zhu et al., 2003). Secondly, when TD is fused to the Gal4-binding domain (GBD), it can promote Sir4p degradation and subsequently break telomeric silencing. In support of its function as a module, the ability of TD to break telomeric silencing is TD-specific; when INC is fused to GBD, no similar effect is observed (Fuest et al., manuscript submitted). Thirdly, Ty5 TD is replaceable. Although we've never tested if Ty5 is still active when TD is removed, replacing TD with exogenous motifs does not compromise IN function. Transposition frequencies of such modified Ty5 elements are comparable to those of wild type elements (data not shown). However, it is quite apparent that protein secondary structure in the vicinity of TD has to be preserved, since not all exogenous motifs can be used to swap TD and still keep Ty5 active (data not shown). Fourthly, and most importantly, Ty5 target specificity can be altered by replacing TD with other peptide motifs that interact with known protein partners. Integration occurs at high efficiency and in close proximity to DNA sites where the protein partners are tethered (Zhu et al., 2003). We speculate that TD works independently of IN catalytic activity and forms a structured domain on the surface of IN, thereby providing easy access to its interacting partner, Sir4p, and to host regulatory factors (discussed below).

Given the evidence that TD is a functional, independent module, we are very curious about its origin. We consider two possibilities by which Ty5 could have obtained this motif. First, TD may have originated spontaneously by random accumulation of mutations in INC after Ty5 invaded Saccharomyces. We view the likelihood of this possibility to be very small. The second possibility is that Ty5 "borrowed" TD from another cellular protein. This process is much easier to envision. Cellular proteins that could have been a source of TD are Sir4p interacting proteins. A screen was carried out to test whether any Sir4C interacting proteins specifically bind to the region on Sir4C where TD interacts. Among several candidates, Esc1p, a nuclear periphery protein (Andrulis et al., 2002), was found to interact

with the same region of Sir4p as TD. Interestingly, when the amino acid (W974) that is critical for Sir4C to interact with TD is mutated to alanine, interaction between Esc1p and Sir4C is also disrupted (Fuerst and Voytas, unpublished data). Whether TD is derived from Esc1p is still under investigation.

The modularity of the Ty5 targeting domain raises another interesting question: are there any other transposable elements that carry similar motifs? If Ty5 obtained a targeting module from a cellular protein, it is very possible that other transposable elements acquired motifs that mediate integration site selection. Sequence alignments of IN from different retrotransposon families identified a chromodomain-like motif in some centromeric retrotransposons in rice and maize (CRR and CRM) (Gao and Voytas, unpublished data). Chromodomains are motifs that interact with modified residues on histones. CRR and CRM are almost exclusively localized at centromeres, suggesting they target to centromeric regions by recognizing specific histone modifications (Jin et al., 2004; Nagaki et al., 2005). It will be of great interest to test whether the chromodomain-like motif in CRR and CRM integrase can function similarly to TD as a targeting determinant.

There are several potential applications that derive from our work on understanding Ty5 targeting mechanisms and identifying Ty5 targeting determinants. The biggest potential application is to develop better gene therapy vectors. Retroviral vectors are widely used for gene delivery in gene therapy, in part because viral integration generates stable, defined, chromosomal insertions (Verma and Somia, 1997). The randomness of retroviral integration, however, is potentially hazardous and could have deleterious genetic effects, for example, by creating loss-of-function mutations or by activating oncogenes. A previous approach to control retroviral integration has been to fuse sequence-specific DNA binding domains to retroviral integrases (Bushman, 1994; Goulaouic and Chow, 1996). This approach has proven effective in *in vitro* integration assays, but because the integrase modifications often compromise viral replication, this approach has not been successfully utilized *in vivo*. The findings described here suggest an alternative approach for controlling retroviral integration, wherein retroviral integrases are modified to carry small peptide motifs that recognize proteins bound to chromosomal target sites. In collaboration with the McCray lab at the

University of Iowa, we are currently modifying feline immunodeficiency virus integrase for restricted, site-specific integration. To date, the results look very promising and we have filed a U.S. patent with the McCray lab. This approach may also make it possible to do site-specific genome manipulation, such as site-specific gene knockouts.

Ty5 integration specificity is regulated by post-translational modification

The targeting domain of Ty5 integrase is post-translationally modified *in vivo*. S1095 phosphorylation is required for productive interaction with Sir4C *in vitro*. This provides direct evidence for the requirement of post-translational modification in Ty5 targeting and reveals one mechanism that cells have adopted to overcome the deleterious effect of transposable element invasion; that is, they control integration specificity by modifying element-encoded proteins.

The specific mechanism used by yeast cells to regulate Ty5 targeting will remain elusive until we identify the kinase(s) responsible for phosphorylating S1095. It is known that Ty5 transposition is regulated by the pheromone response pathway, and Ty5 transcription is induced approximately 20-fold upon pheromone treatment (Ke et al., 1997). Kinases associated with the pheromone response pathway, therefore, are prime candidates for post-translationally regulating Ty5. Extensive studies on Ty1 have revealed that its transposition is also regulated through protein kinase cascades. For example, Ty1 is activated by the Kss1p MAPK cascade (Morillon et al., 2000), which coordinates stress responses (recently reviewed by (Lesage and Todeschini, 2005)); however, there is no evidence that Ty1 integration specificity is regulated by phosphorylation. Because some MAP kinases that sense environmental stress affect Ty5 targeting, this raises the possibility that target specificity may be differentially regulated upon stress. Some of these kinases negatively regulate Ty5 targeting, and therefore upon stress, this may increase the likelihood that Ty5 inserts into other regions of the genome besides heterochromatic domains. This may increase the chance for Ty5 to survive stress conditions. More experimental evidence will be required to further develop this hypothesis.

Another interesting observation we made is that some kinases affect the status of silent chromatin. This suggests that targeting is controlled not only by regulating element-encoded proteins, but perhaps also by affecting the distribution of Ty5's partner, Sir4p. Again, this regulation may be in response to stress, as evidenced by our identification of several kinases involved in stress response pathways that affect the integrity of heterochromatin. Furthermore, we found that the Ty5 integrase C-terminus (INC) is multiplly phosphorylated by different kinases *in vitro*. This *in vitro* data may over-estimate the number of kinases involved in INC phosphorylation, because non-specific phosphorylation can occur in *in vitro* assays. The *in vitro* data must be further validated in future experiments. However, because most of the kinases that phosphorylate TD *in vitro* affect Ty5 targeting *in vivo*, this suggests that the observed IN phosphorylation may have biological relevance.

FUTURE DIRECTIONS

The findings presented in this dissertation provide a starting point for several very interesting future experiments. First, it will be important to identify the kinase that directly modifies the targeting domain. Using the *in vitro* assays and various substrates, we have narrowed down the possibilities to a short list of candidates. This process of elimination should be continued until the kinase that specifically phosphorylates S1095 is identified. The data obtained from *in vitro* assays should also be validated *in vivo*. The identification of the kinase(s) will open up new avenues of investigation for dissecting the specific mechanisms that the host uses to regulate Ty5 targeting.

Another interesting line of experiments will be to test the function of phosphorylation at positions outside of TD. The interaction between TD and Sir4p is weaker than the INC/Sir4p interaction, suggesting that other regions of INC facilitate binding. Phosphorylation at other positions on INC may increase or decrease its affinity for Sir4p. It is also possible that these modified sites may reveal other mechanisms by which Ty5 integration specificity and transposition is regulated. Furthermore, It will be of great interest to test whether phosphorylation of INC can regulate its stability, a function of phosphorylation that has been demonstrated for a number of cellular proteins.

Finally, it will be very interesting to determine how Ty5 overcomes the barrier of heterochromatin to integrate into these regions. Presumably, in order for Ty5 to integrate into the genome, integrase has to bind to naked DNA. However, heterochromatic regions are normally bound with proteins that form a large complex that excludes the access of DNA modifying enzymes. Therefore, there must be a way for Ty5 to open up heterochromatin after integrase interacts with Sir4p. Two possible mechanisms can be envisioned: First, Ty5 integration may be precisely coordinated with cell cycle progression. Therefore, after the Ty5 preintegration complex is tethered to heterochromatin, these regions may be opened up for DNA replication. Ty5 cDNA would then be integrated into the naked DNA before the re-establishment of heterochromatin. This hypothesis is hard to test, but the kinase that modifies TD may give us some hints as to whether or not it occurs or how the hypothesis can be addressed experimentally. The other possibility is that the kinase that modifies TD also regulates the status of heterochromatin. For example, the kinase may phosphorylate some of the Sir proteins, which subsequently leads to their turnover and results in the disruption of heterochromatin. After Ty5 integrates into heterochromatin, silent chromatin is rebuilt. To accomplish this, the kinase that phosphorylates TD and the Sir proteins has to be degraded or excluded from nucleus, or more likely, it has to be inactivated. This process can also be coordinated with cell cycle progression and regulated by cyclin-dependent kinases. Again, the identification of the kinase(s) that modify TD and further characterization of the kinases that affect heterochromatin integrity will be very informative.

REFERENCES

Andrulis, E. D., Zappulla, D. C., Ansari, A., Perrod, S., Laiosa, C. V., Gartenberg, M. R., and Sternglanz, R. (2002). Esc1, a nuclear periphery protein required for Sir4-based plasmid anchoring and partitioning. Mol Cell Biol *22*, 8292-8301.

Boeke, J. D., and Devine, S. E. (1998). Yeast retrotransposons: finding a nice quiet neighborhood. Cell *93*, 1087-1089.

Bushman, F. D. (1994). Tethering human immunodeficiency virus 1 integrase to a DNA site directs integration to nearby sequences. Proc Natl Acad Sci U S A *91*, 9233-9237.

Craig, N. L. (1997). Target site selection in transposition. Annu Rev Biochem *66*, 437-474.

Gai, X., and Voytas, D. F. (1998). A single amino acid change in the yeast retrotransposon Ty5 abolishes targeting to silent chromatin. Mol Cell *1*, 1051-1055.

Goulaouic, H., and Chow, S. A. (1996). Directed integration of viral DNA mediated by fusion proteins consisting of human immunodeficiency virus type 1 integrase and Escherichia coli LexA protein. J Virol *70*, 37-46.

Haren, L., Ton-Hoang, B., and Chandler, M. (1999). Integrating DNA: transposases and retroviral integrases. Annu Rev Microbiol *53*, 245-281.

Jin, W., Melo, J. R., Nagaki, K., Talbert, P. B., Henikoff, S., Dawe, R. K., and Jiang, J. (2004). Maize centromeres: organization and functional adaptation in the genetic background of oat. Plant Cell *16*, 571-581.

Ke, N., Irwin, P. A., and Voytas, D. F. (1997). The pheromone response pathway activates transcription of Ty5 retrotransposons located within silent chromatin of *Saccharomyces cerevisiae*. Embo J *16*, 6272-6280.

Lesage, P., and Todeschini, A. L. (2005). Happy together: the life and times of Ty retrotransposons and their hosts. Cytogenet Genome Res *110*, 70-90.

Morillon, A., Springer, M., and Lesage, P. (2000). Activation of the Kss1 invasive-filamentous growth pathway induces Ty1 transcription and retrotransposition in Saccharomyces cerevisiae. Mol Cell Biol *20*, 5766-5776.

Nagaki, K., Neumann, P., Zhang, D., Ouyang, S., Buell, C. R., Cheng, Z., and Jiang, J. (2005). Structure, divergence, and distribution of the CRR centromeric retrotransposon family in rice. Mol Biol Evol *22*, 845-855.

Peterson-Burch, B. D., and Voytas, D. F. (2002). Genes of the Pseudoviridae (Ty1/copia retrotransposons). Mol Biol Evol *19*, 1832-1845.

Verma, I. M., and Somia, N. (1997). Gene therapy -- promises, problems and prospects. Nature *389*, 239-242.

Xie, W., Gai, X., Zhu, Y., Zappulla, D. C., Sternglanz, R., and Voytas, D. F. (2001). Targeting of the Yeast Ty5 Retrotransposon to Silent Chromatin Is Mediated by Interactions between Integrase and Sir4p. Mol Cell Biol *21*, 6606-6614.

Zhu, Y., Dai, J., Fuerst, P. G., and Voytas, D. F. (2003). Controlling integration specificity of a yeast retrotransposon. Proc Natl Acad Sci U S A *100*, 5891-5895.

Zou, S., Ke, N., Kim, J. M., and Voytas, D. F. (1996). The *Saccharomyces* retrotransposon Ty5 integrates preferentially into regions of silent chromatin at the telomeres and mating loci. Genes Dev *10*, 634-645.

ACKNOWLEDGEMENTS

I would like to gratefully thank Dr. Daniel F. Voytas for his guidance, patience and friendship throughout my time as his graduate student. This dissertation could not have been written without him. Many members of the Voytas lab have influenced me during my time as a graduate student especially Yunxia Zhu and Weiwu Xie. I'm very grateful to Yunxia Zhu for her guidance when I first joined the lab, and to Weiwu Xie for his pioneer work to initiate the project. I would also like to thank Troy Brady for his friendship and a lot of good discussions about science, and Jiquan Gao, with whom I worked closely and tried to puzzle over many of the same problems. I'm also very grateful to the rest of the Voytas lab for making my life extremely enjoyable. These people are, but not limited to, Robert Dick, Fengli Fu, Peter Fuest, Xiang Gao, Yi Hou, Clarice Schmidt, Jeffery Townsend, Kathy Wiederin, Ronnie Winfrey and David Wright. Additionally, I would also like to acknowledge Neil Carroll for helping me prepare some of the figures in this dissertation.

Finally and most importantly, I would like to thank my wife, Fang Peng, for her love and patience throughout my time at Iowa State University. Without her support, it would have been impossible for me to finish this work. One of the best experiences that we lived through in this period was the birth of our lovely daughter Ella Dai, who provided an additional and joyful dimension to our life mission.